The Patent Office

INVESTOR IN PEOPLE

4

EP99/7718

EPO - DG 1

21. 12. 1999

(74)

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

REC'D 29 DEC 1999
WIPO PCT

BEST AVAILABLE COPY

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

I also certify that the attached copy of the request for grant of a Patent (Form 1/77) bears an amendment, effected by this office, following a request by the applicant and agreed to by the Comptroller-General.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

Signed J Evans

Dated 9th December 1999

This Page Blank (uspto)

PCT/EP 99 / 07718

Patents Form 1/77

Patents Act 1977
(Rule 16)

The Patent Office

01SEP98 E386860-1 D02716
P01/7700 25.00 - 9818873.3

THE PATENT OFFICE
28 AUG 1998
RECEIVED BY FAX

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

The Patent Office

Cardiff Road
Newport
Gwent NP9 1RH

1. Your reference

PAT : 98101 GB

2. Patent application number
(The Patent Office will fill in this part)

9818873.3

3. Full name, address and postcode of the or of each applicant *(underline all surnames)*

NOKIA OY
KEILALAHDENTIE 4
02150 ESPOO
FINLAND

Patents ADP number *(if you know it)*

07504434001

If the applicant is a corporate body, give the country/state of its incorporation

FINLAND

4. Title of the invention

A METHOD AND SYSTEM FOR SUPPORTING THE QUALITY OF SERVICE IN WIRELESS NETWORKS

5. Name of your agent *(if you have one)*

"Address for service" in the United Kingdom to which all correspondence should be sent *(including the postcode)*

HELEN LOUISE HAWS
NOKIA MOBILE PHONES
PATENT DEPARTMENT
ST GEORGES COURT
ST GEORGES ROAD
CAMBERLEY
SURREY GU15 3QZ UK

Nokia UK Ltd
Nokia IPR Department
Nokia House, Summit Avenue
Southwood, Farnborough,
Hampshire GU14 0NZ

Patents ADP number *(if you know it)*

06945539001

EGP. 26.1.99.

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and *(if you know it)* the or each application number

Country	Priority application number *(if you know it)*	Date of filing *(day / month / year)*

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

Number of earlier application	Date of filing *(day / month / year)*

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? *(Answer 'Yes' if:*

a) any applicant named in part 3 is not an inventor, or YES
b) there is an inventor who is not named as an applicant, or
c) any named applicant is a corporate body.
See note (d))

Patents Form 1/77

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form	
Description	63 + 36p Annex 1
Claim(s)	4
Abstract	1
Drawing(s)	13

10. If you are also filing any of the following, state how many against each item.

Priority documents	
Translations of priority documents	
Statement of inventorship and right to grant of a patent *(Patents Form 7/77)*	
Request for preliminary examination and search *(Patents Form 9/77)*	1 ✓
Request for substantive examination *(Patents Form 10/77)*	
Any other documents *(please specify)*	

11. I/We request the grant of a patent on the basis of this application.

Signature H L HAWS Date 28.8.1998

H L HAWS - Agent for the Applicant

12. Name and daytime telephone number of person to contact in the United Kingdom

Kendra Jeffery - 01276 419538

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

a) *If you need help to fill in this form or you have any questions, please contact the Patent Office on 0645 500505.*

b) *Write your answers in capital letters using black ink or you may type them.*

c) *If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.*

d) *If you have answered 'Yes' Patents Form 7/77 will need to be filed.*

e) *Once you have filled in the form you must remember to sign and date it.*

f) *For details of the fee and ways to pay please contact the Patent Office.*

DUPLICATE

A method and system for supporting the quality of service in wireless networks

The present invention relates to a mechanism for supporting the quality of service in radio networks. In particular, it relates to a method, system and communication devices which support the quality of service in packet data transmission in a radio network.

Such a mechanism is required, for example, in wireless internet protocol (IP) networks.

The term "Internet" is commonly used to describe an information resource from which information can be retrieved from a data processor, such as a personal computer (PC). The data processor communicates via a modem with a telecommunication network. This information resource is distributed worldwide, comprising several storage locations which also communicate with the telecommunication network. The Internet is made operable by defining certain data communication standards and protocols, such as TCP (transfer control protocol), UPD (user datagram protocol), and IP (Internet protocol), which are used for controlling data transmission between numerous parts of the Internet. The TCP and the UDP are involved with preventing and correcting data transmission errors in the data transmitted in the Internet; the IP is involved with data structure and routing. The currently used versions of the Internet protocol are IPv4 and IPv6. IPv4 is defined in RFC791 and IPv6 is defined in RFC1883.

Thanks to the growing popularity of open data systems, the Transmission Control Protocol/Internet Protocol (TCP/IP) communication protocol has become a generally used protocol whereby computers of different sizes and brands can communicate with each other. TCP/IP support is currently available for almost all operating systems. The network layer protocol of TCP/IP, the Internet Protocol IP, is intended to be routed by gateways, *i.e.* routers. The routing is conducted by means of IP addresses of four bytes and routing tables. Thanks to the Internet protocol, computers using the TCP/IP can transfer messages in the routing network even to the other side of the world.

The rapid evolution of the Internet services has created a strong need for broadband networks with high data rate and Quality of Service (QoS). Video broadcasting and other multimedia distribution services are evolving rapidly. The users are willing to access these services also in the wireless environment. Currently, in the fixed IP network, IP packets are typically sent as best effort data traffic. In the case of network congestion, all data streams are handled with equal priority which may have a dramatic effect on multimedia services. Two main problems exist: firstly the current wireless networks will not provide sufficient QoS mechanism, and secondly, the existing wireless networks are not capable of serving several simultaneous connections with high data rate and QoS requirements. To meet the increased customer requirements, new wireless broadband network techniques are required.

The Internet Engineering Task Force (IETF) is an organisation involved with the development of internet architecture and operation in the internet. They have defined two different QoS concepts: integrated and differentiated services, for providing a standard mechanism for supporting real time applications in IP networks. Integrated services is based on an abstract flow

model with reservation protocol (RSVP, RFC2205) and admission control. The network reserves statically resources for real time connections in each network device, and hence is not optimally efficient. Consequently the differentiated services concept was developed. This concept is based on the use of an IP header for indicating the requested service class for the packet. As a result, each IP packet carries QoS information and no reservations are required. Whilst the IETF suggest the use of an IP header for indicating the QoS, the actual packet handling mechanisms will not be standardised.

The present invention provides a wireless IP network architecture which supports QoS, and in particular differentiated services.

According to a first aspect of the present invention, there is provided a method for supporting the quality of service in packet data transmission in a radio network, whereby transmission over the air interface is in radio flows, the method comprising selecting a radio flow having appropriate quality of service characteristics for the packet to be transmitted over the air interface from a selection of predefined default radio flows having different quality of service characteristics.

The provision of default radio flows enables the radio network to support differentiated services, and thus cater for wireless broadband services. In particular, it prevents bottlenecking of data in the radio part of an IP network.

Radio flow selection may be effected by providing the packet to be transmitted with a radio flow identifier selected from predefined default radio flow identifiers representative of different quality of service characteristics. The packet may then be mapped into the identified default radio flow for transmission over the air interface.

Advantageously, the method comprises detecting handover of a mobile communications device having an active connection from one radio

subnetwork to another and performing default radio flow selection for the active connection in response to handover detection.

This further improves inter-domain mobility. For example, as a mobile terminal moves into a new radio subnetwork (e.g. the area of a new mobile router), the default radio flows can be utilised for improving the performance of the handover. The new router will typically have no information of the terminal connections and requested QoS. Therefore, it cannot establish QoS flows in advance. In order to reduce packet losses during the handover, the terminal and network can temporarily switch the existing connections ("old flows") into one of the default radio flows. This approach allows the assignment of a higher QoS for certain connections until the new router detects QoS (IP) flows and switches the connections into separate radio flows.

Preferably, the method also supports integrated services. For example, in a preferred embodiment of the invention, the method further comprises monitoring packets to be transmitted over the air interface to detect IP flows, and switching a detected IP flow to a dedicated radio flow having corresponding quality of service characteristics. Typically, a default radio flow will initially be selected, and then an IP flow will be transferred to a separate radio flow once that radio flow is set up.

Switching the detected IP flow to a dedicated radio flow may be effected by providing the packets of a detected IP flow with an identifier of the dedicated radio flow, and mapping the packets of the detected IP flow into the identified dedicated radio flow for transmission over the air interface.

A system and communication devices are provided which implement the aforementioned method.

For example, according to another aspect of the present invention, there is provided a radio access system for supporting the quality of service in data packet transmission over the air interface, the system comprising a selection of predefined default radio flows having different quality of service characteristics, and means for selecting a radio flow having appropriate quality of service characteristics for the packet to be transmitted over the air interface from the selection.

The radio flow selecting means optionally comprises means for providing the packet to be transmitted with a radio flow identifier selected from identifiers corresponding to the predefined default radio flows and means for mapping the packet into the identified default radio flow for transmission over the air interface.

Also, the system preferably comprises means for detecting handover of a mobile communications device having an active connection from one radio subnetwork to another. In this event, the selection means selects a default radio flow for the active connection in response to handover detection.

Advantageously, the system further comprises means for monitoring packets to be transmitted over the air interface to detect IP flows, and means for switching a detected IP flow to a dedicated radio flow having corresponding quality of service characteristics. The switching means may comprise means for providing the packets of a detected IP flow with an identifier of the dedicated radio flow, and means for mapping the packets of the detected IP flow into the identified dedicated radio flow for transmission over the air interface.

According to a further aspect of the present invention, there is provided a communication device for use in a system which supports the quality of service in data packet transmission over the air interface and comprises a selection of predefined default radio flows having different quality of service characteristics, wherein the device is arranged to select a default radio flow having appropriate quality of service characteristics for the packet to be transmitted over the air interface from the selection. The communication device may, for example be a mobile communication device, an access point controller or a mobile router.

Embodiments of the present invention will now be described, by way of example, with reference to the accompanying drawings, of which:

Figure 1 illustrates a GRAN Reference Model;

Figure 2 illustrates a General System Architecture;

Figure 3 shows target operating environment and transmission link alternatives;

Figure 4 gives an overview of the data path architecture;

Figure 5 illustrates transmission link alternatives;

Figure 6 shows an Ethernet based system architecture;

Figure 7 shows the data plane for ATM;

Figure 8 shows the data plane for Ethernet;

Figure 9 illustrates an exemplary structure of a radio access network identifier (RAN_ID);

Figure 10 illustrates packet encapsulation if an ATM transmission link is used;

Figure 11 illustrates flow mapping in the case of an Ethernet transmission link;

Figure 12 outlines of the defined flow management scheme;

Figures 13 and 14 illustrate an IP and radio flow multiplexing scheme ;

Figure 15 illutrates the system architecture and main interfaces;

Figure 16 fives an example of different flow types;

Figure 17 illustrates typical WFMP tables;

Figure 18 shows mobile router (M-router) WFMP routing for downlink packets;

Figure 19 shows M-router WFMP routing for uplink packets;

Figure 20 illustrates transmission between two mobile terminals;

Figure 21 illustrates the mobile terminal side WFMP operation for the uplink;

Figure 22 illustrates the mobile terminal side routing for the downlink;

Figure 23 illustrates a flow compression (FC) structure;

Figure 24 illustrates Integrated Services;

Figure 25 illustrates Differentiated services;

Figure 26 illustrates Port information;

Figure 27 illustrates H.323; and

Figure 28 shows a Wireless QoS driven queuing and error control strategy.

A wireless IP network architecture of a preferred embodiment of the invention will now be described, which enables efficient mobile access to both narrow-band packet data services and to delay-sensitive broadband multimedia services. Firstly, the system architecture will be described, including the

network structure, data plane architecture, and IP flow management and architecture.

Firstly, the general system architecture defining the main physical blocks, protocol framework and the main internal interfaces will be discussed. Then, the functionality and interworking of defined functional entities and protocols will be illustrated.

Network Structure

Theoretical Reference model

The system architecture follows the outlines of the theoretical General Radio Access Network (GRAN) reference model, illustrated in Figure 1 of the accompanying drawings [ES21]. A broadband radio access network 1 (BRAN) includes a radio access network 2 (RAN) and core network dependent interworking functionality (IWF) blocks. The RAN holds all the radio dependent parts and the IWFs link RAN to various core networks 3 and terminal entities 4. Hereafter in the description, the term preferred system generally refers to the entire BRAN network covering RAN and necessary IWFs. The preferred system is connected to the IP core network.

Network Entities

The broadband radio access network depicted in Figure 2 is composed of a radio access network 2 having mobile terminals 4, access points 51, 51' and an air interface between, plus a mobility enhanced IP router 52(M-Router). The BRAN is connected to the core IP network which comprises the internet backbone 21 and home agents 22.

The radio access network 2 (RAN) implements all the radio dependent functionality such as radio resource management, setup and release of wireless flows, handovers and packet compression. It contains mobile

terminals and access points. The mobile terminal 4 is the user's communication device for accessing wireless Internet services, and is the end point of the Internet and radio access network control protocols. The access point 51,51' implements all the radio dependent control functionality, such as radio resource management. It includes radio resource management and radio link control functions. The corresponding network elements in GSM are the base transceiver stations (BTS/TRX) and base station controller (BSC).

The M-Router 52 creates the wireless IP sub-network managing one or more access points. The M-Router handles the mobility and location management of the terminals that are registered to the access points 51,51'. The M-Router provides IP mobility services, such as DHCP (dynamic host configuration protocol). DHCP is used for allocating IP addresses for the terminals. Corresponding element in the GSM network is the gateway GPRS support node (GGSN). The access points 51, 51' and the terminals 4 with an IP stack that belong to the same IP sub-network (use the same M-ROUTER) create a logical link.

The core network 3 comprises a home agent 22 which resides in the home network of an associated terminal 4 and is accessed through standard IP gateways. Typically home agent 22 is implemented as part of the M-Router 52 of the home network. However, it can also be a separate entity (e.g. PC host). The home agent 22 can contain user authentication information and a billing data base. It resembles the home location register (HLR) in GSM.

Target Environment

This preferred system of the invention may be used for private and public networks. Public networks are typically operated by Internet service providers or telecom operators. Common places for public wireless access network are hot spots, such as airports, hotels, railway stations etc. In this case the public

network operator has to be able to reliably authenticate the users for billing purposes. In addition the network should offer security on IP level.

Business local area networks (LANs) provide another interesting application area for the system of the present invention. Here the system provides a wireless extension for the existing fixed LAN infrastructure. Typical company LANs are based on Ethernet cabling. Therefore, in such an implementation the access points may be connected to the M-Router 52 via Ethernet. The M-Router can serve both fixed terminals and the mobile terminals 4. Figure 3 of the accompanying drawings illustrates various transmission link alternatives.

Data Plane

The general architecture of the system at the OSI data layer level will now be described.

In the preferred embodiment of the invention, the M-Router 52 has a full TCP/IP stack functionality. It performs standard IP routing forwarding packets to the RAN interface and embeds wireless specific control functions. As will be explained further below, in accordance with the invention, the M-Router 52 classifies the incoming IP packet flows and relays them via the corresponding access point to the mobile terminal using the suitable QoS characteristics. The wireless extension controls radio flows, terminal mobility and location management. The M-Router 52 controls the access points using a specific control protocol. The access point 5 implements a LAN bridge that multiplexes the IP flows into radio flows.

The mobile terminal 4 includes all standard TCP/IP entities and wireless specific control services. The control messages are transparently sent between the M-Router 52 and terminals 4 utilising control functions. Figure 4 illustrates the data path architecture.

The mobile router 52 segments and re-assembles IP packets into segments that fit into radio link packets. The segmentation and re-assembly (SAR) blocks handle the segmentation between the mobile terminal 4 and the M-Router. The access point only transparently relays the segmented packets between the radio access network 2 and the fixed network. BRAN radio utilises asynchronous transfer mode (ATM) like segmentation (AAL), which segments the IP packets into 48-bit cells.

Transmission Link Alternatives

The system offers two physical transmission links for connecting access points to the mobile router: ATM and Ethernet. Here ATM refers to the physical link without control signalling. In this case ATM provides only segmentation and re-assembly and transmission services for IP traffic. The ATM option is intended for public telecommunication networks owned by a public operator whilst, as mentioned above, the Ethernet offers an ideal solution for private wireless business LANs. It is assumed that public hot spots are built as separate systems, which offers a good opportunity to deploy ATM links for connecting mobile routers to the backbone networks. Figure 5 illustrates these two transmission link options.

In the Ethernet case all access points belong to the same Ethernet segment which is connected to the router e.g. through a hub. The access points in the same Ethernet segment have the same IP sub-network address. Therefore, they comprise one IP sub-network (called a mobility domain). The same router can control fixed network elements that are located in a separate IP domain. It is even possible to install access points to different router ports and to create several mobility domains with different IP network address. Figure 6 illustrates the resulting architecture.

Figure 7 of the accompanying drawings illustrates packet encapsulation in the ATM case. In the ATM link the M-Router port identifies the access point

uniquely. The IP packets are encapsulated into AAL5 frames between the M-Router and the mobile terminals. AAL5 layer segments the packets into 48-bit long ATM cell payload. In this case the access point passes the ATM cells transparently to the M-Router.

Figure 8 of the accompanying drawings illustrates packet encapsulation in the Ethernet case. In the Ethernet link IEEE MAC addresses are used for identifying access points and the M-Router ports. The M-Router and mobile terminals encapsulate IP packets inside Ethernet frames. Before packets are passed into Ethernet layer (SAR) they are passed through an IP/Ethernet convergence layer. This convergence layer adds a dedicated flow label between the IP packet and the Ethernet header and its concept is the subject of a copending patent application FI980191. (A copy of this application is attached hereto as Annex 1 and its content is incorporated into the present application by reference). This flow label corresponds to ATM virtual path identifier/ virtual connection identifier (VPI/VCI) values. It is decoded at the access point (bridge) which multiplexes Ethernet packets into correct radio flows. The mapping between radio flows and Ethernet frames is done in the convergence layer.

In both cases the convergence layer marks the IP packets with radio access network specific RAN identifier (RAN_ID). In the case of ATM the RAN_ID corresponds to the VPI/VCI value while the Ethernet uses a random 24-bit identifier allocated by the M-Router. The first 8-bits of the RAN_ID are used for identifying the terminal and the rest 16-bits for identifying the connection. The RAN_ID is unique within the access point. The ATM uses port to detect access point while the Ethernet case utilises AP IEEE MAC addresses. Figure 9 illustrates the structure of the RAN_ID and the mapping into the radio flows according to this preferred embodiment of the present invention.

If ATM transmission link is used the flow mapping and the use of the RAN_ID (VPI/VCI) is trivial, and Figure 10 illustrates the packet encapsulation for this case.

In the Ethernet case the RAN_ID can be any random 24-bit identifier which is added between the IP packet and Ethernet headers. However, in order to achieve a scalable system specification the allocation of 24 bits in Ethernet case is preferably selected so that it corresponds to the ATM case. That is, so that the first 8 bits identity the terminal and the next 16 bits identify the radio flow. As a result, the encapsulation of Ethernet packets is compatible with ATM case, except that the segment size differs (ATM has 48-bits). A dedicated protocol ID has to be defined and added in the Ethernet header for indicating the existence of RAN_ID in front of IP packet. Figure 11 illustrates the resulting Ethernet packet.

Outline of IP Flow Management

The internet protocol implements a connectionless packet data system. The data is carried inside packets, the header of which indicates the correct destination address. This transmission scheme does not enable the system to separate various connections. The only possibility to detect connections is to monitor the IP traffic inside the M-Router 52 and to try to detect and classify IP packet streams, called IP flows. An IP flow is established if two hosts (applications) frequently (or regularly) send IP packets between each other. Flow classification is explained in more detail below. The network can assign certain QoS characteristics for a flow, which is required for multimedia service implementation in IP networks. For instance a particular flow can be prioritised in the router.

The mechanism of the invention maintains IP flow QoS characteristics in the air interface and permits the prioritisation of different IP packets (flows) in the radio link. The concept deploys radio flows which are created between the

mobile terminal and the access points. The outlines of the specified flow management in the preferred embodiment are as follows:

The M-Router 52 monitors the headers of the incoming IP packets and tries to classify the existing IP flows (i.e. regular packet streams) utilising the IPv6 flow label and/or destination and source IP addresses and ports. If the M-Router detects an IP flow, it will start marking the packets which belong to this flow with a specific RAN_ID (Radio Access Network IDentifier). The router allocates a unique RAN_ID for each detected flow. RAN_ID is utilised for separating packets belonging to IP flows in the radio access network. Consequently, the use of RAN_ID will create "virtual connections" through which IP flows are packed cross the RAN (MT-M-ROUTER). RAN_ID corresponds to the ATM VPI/VCI identifiers. A wireless flow management protocol (WFMP) is disclosed in the aforementioned copending application (Annex 1) for managing RAN identifiers. Both the terminal 4 and the M-Router 52 have WFMP entities which communicate peer-to-peer over the wireless link. WFMP actually provides the convergence layer functionality. The M-Router WFMP always detects flows, allocates RAN_IDs and informs mobile terminal WFMP of the assigned ID value. To minimise the overhead the 24-bit RAN_ID is compressed into 8-bit MVC (Mobile virtual circuit) identifier in the radio link. Accordingly, the MVC scheme supports 256 flows per terminal.

As mentioned previously, the radio link is the bottleneck concerning QoS and the throughput. The present invention addresses this problem in this preferred embodiment by providing a system which includes two mechanisms which improve the wireless support for broadband services: priority queues and flow compression. The radio sub-system handles various radio flows differently. It has three separate buffering queues for the incoming traffic: high priority queue for real-time traffic, medium priority queue for non-realtime data and low priority queue for best-effort data. Normally the M-Router establishes flows, but the system enables also the mobile terminal to request the M-

Router to establish a radio flow (RAN ID) with a given priority for a particular IP flow. As the M-Router classifies the IP flow it assigns one of the three priority classes for the established radio flow. The decision is made on the basis of the IP flow type and traffic characteristics. Various radio flow priorities enable wireless broadband services.

IPv6 headers are pretty long (e.g. IPv6 address is 128 bits). The radio overhead is minimised by compressing the IP headers of detected flows. The compression is performed between the M-Router and the mobile terminal. For this purpose these entities include specific Flow Compression (FC) entities. The IP header compression is efficient for flows as the IP source and destination can be identified also from RAN_ID. The receiving end can look up the RAN_ID and add the missing parts of the IP header on the basis of it.

Only the detected IP level flows will be switched into separate radio flows. The other IP packets are transmitted over the air in default channels. The system of this embodiment defines three fixed (hard-coded) radio flow identifiers which are used for transmitting non-flow packets. In this case the IP packets from multiple different sources are multiplexed into the same radio flow (RAN_ID), which will make IP header compression impossible. However, the system offers three multiplexed radio flows per terminal, one for each radio priority queue. As mentioned above, these "default" radio flows are provided according to the present invention for improving the support for inter-IP domain mobility and differentiated services. In the scheme of the preferred embodiment, the M-Router 52 or the terminal 4 can look at the priority bits of a single IP packet and send that with the corresponding radio QoS (priority). Then, as soon as an IP flow is detected, the packets can be switched into a separate radio flow with certain QoS and with IP header compression.

Figure 12 summarises the defined flow management concept for IP flows. The detected IP flows are marked with a dedicated RAN_ID label (in ATM VPI/VCI). The access point performs RAN flow – radio flow multiplexing, i.e.

mapping between RAN_ID and MVC. The M-Router manages the location of MTs and updates the routing table if the mobile performs handover between access points.

IP Flow Multiplexing

Figures 13 and 14 illustrate the flow management procedure for downlink traffic. The process is as follows:

The M-ROUTER monitors the incoming traffic continuously. As the amount of packets per IP flow (between certain hosts (ports)) exceeds the threshold value (per time), the WFMP establishes a RAN flow and allocates new RAN_ID for it. Next the packets belonging to the flow are passed to the access point via Flow Compression (FC) using the allocated flow specific RAN_ID. In accordance with the invention, the IP packets that do not belong to any flow are marked with one of three default RAN_IDs for providing differentiation.

The FC entity compresses the IP header of the detected flows and copies RAN_ID to the resulting packet. In the receiving end the peer FC entity can detect the correct source by decoding RAN_ID and assigns the missing IP header before the packet is passed to the upper layers. Only detected IP flows are compressed.

In this embodiment, the access point 51, 51' has a conversion table which maps the MVCs (radio flows) into correct RAN_IDs. The M-Router 52 allocates RAN_ID address space per access point. The packets are then transmitted to the M-Router with RAN_ID "flows". Next, the packets are transmitted to the radio link. The default RAN_IDs are assigned with a fixed radio link priority. In this embodiment, three default "pipes" (RAN_IDs) exist for user data: real-time flow, non-real time flow and best-effort flow. Of course, an alternative number of default channels could be used depending on requirements. For example, two channels – high and low priority might be

sufficient in some instances, where congestion is not likely to be such a problem, and in other circumstances, the separation into a further number of QoS characteristics might be required. Each channel or pipe has a pre-configured RAN_ID. The non-flow IP traffic is transmitted within these flows without any compression.

The radio of the access point (layers 1 and 2) allocates unique MVC values per RAN_ID. Three default MVC "pipes" exists for the non-compressed traffic also in the air interface. The default RAN_IDs are mapped into the corresponding hard-coded MVCs while compressed IP flows are switched into dedicated MVC connections. Different flows are separated in the in the air interface using MVC and terminal wireless MAC addresses.

In the mobile terminal 4 of this embodiment, its radio modem converts the received MVC into the corresponding RAN_ID value and passes the packet to the SAR layer which reassembles the data into IP packets still maintaining RAN_ID information. The compressed flows are then passed to FC which identifies the RAN_ID and adds the correct IP header information. The default RAN_ID traffic is passed directly to the WFMP.

Protocol Architecture

The detailed system architecture of the preferred embodiment of the present invention is illustrated in Figure 15. The main external interfaces are listed in the following tables.

Table 1: External Control Interfaces

IF#	Interface	Explanation
1.	MMC – MMC	Mobility management messages between terminal and M-Router. Mobility management messaging is used as a new terminal registers into the network and in the case of handovers.
2.	WFMP – WFMP	Flow management control signalling messages. This is used for establishing and releasing radio flows.
3.	MCP – MCP	Mobile Control Protocol (MCP) provides a reliable peer-to-peer protocol for transmitting WFMP and MMC messages between the mobile terminal and the M-Router. MCP is used for all wireless specific signalling.
4.	Radio control messages	Radio control messages are used for transmitting radio link control messages. For instance terminal association and radio flow (MVC level) control signalling is carried here.
5.	APCP interface	Access Point Control Protocol (APCP) is used for sending radio link control and radio resource management messages between the access points and the mobile router

The external control interfaces define the logical interface between the mobile terminal 4 and the radio access network (access points 51, 51', M-Router 52) and between the radio access network 2 and the core network 3(AP-M-

Router). The external interfaces have to be standardised, if the target is to define compatible standard system which can be composed of devices from different manufacturers.

In addition to the standard interfaces the system has also several important internal control interfaces, which are listed in the table below:

Table 2: Internal Control Interfaces

IF#	Interface	Explanation
6.	Wireless QoS controller – WFMP interface	This is an internal interface which is used for transmitting flow establishment requests and QoS information between the QoS manager and the WFMP. As M-ROUTER-WFMP detects a new flow it queries the radio link priority from the QoS manager. In a real implementation QoS manager and WFMP can be integrated into a single entity which removes interface 6.
7.	Wireless QoS controller – H.323 interface	This interface is used for relaying explicit H.323 traffic characters, such as delay and throughput to the radio link QoS. H.323 control signalling is decoded in the mobile terminal and the required QoS is communicated to the proprietary wireless QoS manager. QoS

		manager will use manufacturer specific rules for converting H.323 QoS into radio flow priority. NOTE: H.323 specification defines a standard interface to the RSVP. This could be also used here instead of implementing dedicated connection between QoS and H.323.
8.	Wireless QoS controller – RSVP interface	Wireless specific QoS controller interacts with RSVP module for obtaining resource reservations and converting those into radio resource reservations and radio QoS. The RSVP module requests certain QoS from wireless QoS manager via this interface.
9.	Wireless QoS controller – MMC interface	This interface is used in the case of handovers. The MMC module informs the QoS controller about the handover and requests QoS controller to re-establish wireless flows. This establishment request is then forwarded to WFMP entity.

The system architecture includes the following functional blocks specific to the preferred embodiment of the invention:

QoS - Wireless QoS controller: This entity allocates the radio link QoS for the IP packets. QoS controller has an interface to H.323 and RSVP blocks which can give explicit QoS requirements, such as delay, bandwidth, for the IP flow. If no explicit QoS parameters are available QoS manager assigns the QoS on the basis of DS field (differentiated services) or on the basis of port information (standard applications). As WFMP detects a flow it informs M-Router-QoS entity of the evaluated packet throughput, this information can be also deployed for allocating radio link QoS. The QoS controller transmits the allocated radio link QoS values to the WFMP entity which then establishes radio flows with selected QoS. The QoS controller can also send QoS update messages through WFMP. Using QoS update message QoS controller can change the radio QoS value for existing flows. This functionality is useful for instance if H.323 or RSVP parameters change during the connection.

Mobility Management Controller (MMC): MMC entity is responsible for the terminal mobility management. The M-ROUTER-MMC has a database which contains information of the registered terminals and their current location (access point). During the terminal registration MMC can be used for authenticating the user. The mobile terminal MMC initialises the handover by sending a handover_request message to the M-Router-MMC which checks the radio resources in the new access point and requests WFMP to establish new radio flows in the new access point and to release the old radio flows.

Wireless Flow Management Protocol (WFMP): WFMP entity manages the radio flows. It detects IP traffic and classifies IP flows. As WFMP detects a new flow it passed flow information to QoS controller that assigns the correct radio link priority for the flow. Next WFMP establishes the radio flow with the allocated priority. WFMP allocates RAN_IDs and updates access point RAN_ID – radio flow tables. The M-Router includes a master WFMP which

classifies the flows and maintains the data base of all the existing flows while the mobile terminal includes only a simple WFMP entity that multiplexes the RAD_IDs into correct radio flows. MT-WFMP can request the M-ROUTER-WFMP to establish a new radio flow with a given priority. This is the case for instance if MT-H.323 requests a flow with explicit QoS. Also the inter-router handover procedure can utilise this mechanism for quickly setting up flows between the mobile terminal and the new M-Router after the handover.

Mobile Control Protocol (MCP): MCP protocol transmits WFMP and MMC messages between the mobile terminals and the M-Router. MMC provides a reliable mechanism for transmitting control information. MMC implements a simple go-back-N type of retransmission protocol. A separate low layer protocol was added instead of using TCP/IP in order to guarantee a reliable transmission of control messages. TCP/IP will not allow separating control messages from the other TCP traffic. Therefore, e.g. in the case of handover the control messages would be mixed with the user data traffic, which causes a significant delay for the handover procedure and re-establishing connections. The use of MCP allows to prioritise all the control traffic before the user data packets.

Flow Compression (FC) block: The M-Router and mobile terminals include FC entities which compress the detected IP flows. Flow compression is used only for the classified IP flows. The other IP traffic is sent without compression.

Radio Resource Manager (RRM): Each access point has a RRM entity that manages the radio resources of the particular access point. In the present system WFMP sends resource queries to RRM each time a new flow is established. WFMP transmits the requested radio flow priority (allocated by QoS) and estimated flow rate (WFMP evaluator). Based on this information RRM decides whether the connection is accepted or not. RRM provides a mechanism for each access point to dynamically manage the radio resources. RRM compares the estimated flow rate to the free capacity of the requested

radio flow. If insufficient capacity is available RRM can propose WFMP to use lower QoS. If WFMP gets a return QoS value which is lower than the requested QoS it can either stop establishing the connection or to continue with lower QoS. In the case of IP RRM capacity calculations are mainly based on the estimated flow rate (M-ROUTER). Only RSVP and H.323 signalling provide explicit throughput and delay requirements which can be directly mapped into RRM.

Access Point Control Protocol (APCP): APCP protocol provides a mechanism for transmitting control messages between the access points and the M-Router. ACPC can be located on top of TCP/IP stack which guarantees a reliable transmission of control messages. WFMP deploys ACPC for RRM queries and for sending flow control information to the radio sub-system.

The detailed functionality of the system is explained below.

IP Flow Management

As mentioned above, the system of the present invention uses IP flow detection idea for managing connections in the radio access network. The idea is to separate IP flows from the best-effort IP traffic to be able to 'treat' the flows differently. The different treatment means that different priorities (QoS classes) can be given to the detected flows, and that the detected flows can be compressed to save the radio resources.

The present invention resides in the provision of default channels for at least two different QoS classes. In this embodiment, each mobile terminal has three default radio channels: one for best-effort traffic, one for medium priority traffic and one for high priority traffic. All the default channels are terminated at the M-Router WFMP module (n-to-1 relationship). In addition to the default channels, mobile terminals can have detected flows. If the other end of a

detected flow is located at the IP core network, the flow is terminated at the M-Router compression module (all flows are compressed). In case the correspondent node is another mobile terminal, the flow is switched directly between the AP interfaces. Figure 16 shows how the different flows are routed inside the network. The figure presents a simplified network architecture where two mobile terminals are connected to the M-Router through two access points (only the AP interfaces are shown). Each MT has the default channels (three of them), one flow to the IP network, and one flow switched between the mobile terminals.

Next, the WFMP flow management functions are described in detail. Both the M-Router and the mobile terminal side operations are explained.

Flow classification

WFMP is responsible for flow detection and routing of IP packets between the network and mobile terminals/radio connections. The function handling flow detection is called the *flow classifier*. Flow classifier decides when IP packets belong to a flow and after the decision binds the flow to a radio connection. It also needs to detect when a flow terminates and release the corresponding radio connection.

Flow classification works so that WFMP monitors the IP traffic and specific header fields in order to detect new flows. Depending on the IP and transport protocol header fields, the following four different flow types can be specified:

1. Flows identified by flow labels (type 1)
2. Flows identified by TCP/UDP port numbers (type 2)
3. Flows identified by the source and destination IP addresses + the security parameter index (type 3)
4. Flows identified by the source and destination IP addresses (type 4)

The first option can be applied if the applications are able to use the IPv6 flow label to mark the different IP sessions. If such advanced applications are not available, and if TCP/UDP port information is available, flow type 2 is selected. In case IP encryption is used, the second option cannot be applied since the port information is encrypted. In such a case, security parameter index (SPI) is used with source and destination addresses to identify possible flows. If no TCP/UDP port information, flow labels or SPI parameters are available, the only option is to look just for the source and destination IP addresses and separate flows between hosts (the first two options separate flows between IP sessions/processes).

Each flow type specifies the set of fields from the IP packet header that are used to identify a flow. The set of the header fields identifying a particular flow is called the flow identifier. Depending on the flow type, the flow identifiers contain the following fields:

- Type 1: source address + destination address + flow label
- Type 2: source address + destination address + protocol (next header) + source port + destination port
- Type 3: source address + destination address + security parameter index (SPI)
- Type 4: source address + destination address

WFMP can differentiate between these four flow types, and based on the flow classification mechanism bind each flow type to a flow. Three different flow classifier mechanisms which may be applied in the present system are:

- X/Y classifier, meaning X packets (with the same flow identifier) in Y seconds resulting in a new flow
- Protocol classifier which simply assigns all TCP packets to flows

- Port classifier, using transport layer port numbers to decide which flows to bind.

The X/Y classifier is the preferred choice as it is the only one which supports flow types 1 and 2

Flow detection criteria

Typical flow detection criteria for the X/Y classifier are listed in Table 3 X/Y classifier recommendations. The table gives values for X and Y in a function of different amount of flow space available (in this case the flow space refers to the amount of radio connections required). Expected performance means the portion of packets switched to flows.

As can be seen, the values are somewhat different in different environments. Therefore, it should be possible to change easily the values of X and Y in the WFMP implementation.

Table 3: X/Y classifier recommendations

Flow space req.	Gateway	Campus/Enterprise Backbone
1K	Classifier: X = 5 / Y = 15 sec. Flow deletion delay: 30-120 sec. Expected performance: 85%	Classifier: X = 40 / Y = 40 sec. Flow deletion delay: 30-60 sec. Expected performance: 79%
2K	Classifier: X = 5 / Y = 60 sec. Flow deletion delay: 30-120 sec. Expected performance: 90%	Classifier: X = 10 / Y = 45 sec. Flow deletion delay: 30-60 sec. Expected performance: 89%
8K	Classifier: X = 2 / Y = 60 sec. Flow deletion delay: 30-120 sec. Expected performance: 93%	Classifier: X = 5 / Y = 60 sec. Flow deletion delay: 30-60 sec. Expected performance: 92%
16K	Classifier: X = 2 / Y = 60 sec. Flow deletion delay: 30-120 sec. Expected performance: 93%	Classifier: X = 2 / Y = 60 sec. Flow deletion delay: 30-60 sec. Expected performance: 95%
32K	Classifier: X = 2 / Y = 60 sec. Flow deletion delay: 30-120 sec. Expected performance: 93%	Classifier: X = 2 / Y = 60 sec. Flow deletion delay: 30-60 sec. Expected performance: 95%
∞	Classifier: all packets Flow deletion delay: ∞ Expected performance: 99%	Classifier: all packets Flow deletion delay: ∞ Expected performance: 98%

Since the establishment of a TCP connection always contains at least three packets used, and since the flow detection should be based on actual data packets, a minimum value of six for X is considered appropriate (third data packet triggering the flow detection). The value for Y could be 30 seconds.

Flow deletion

A flow is deleted after some constant number of seconds of inactivity. When flow classifier detects a new flow, it starts the flow inactivity timer. This timer is re-started each time a packet belonging to that flow is received. Once the timer expires, the flow identifier is removed from the list of monitored packets. Finally, WFMP releases the flow both from the RAN and the mobile terminal.

RSVP reserved flows

When RSVP is used to reserve resources from the network, the M-Router RSVP module needs to communicate with the mobility specific modules to reserve resources from the radio access network. RSVP module has an interface to the QoS manager and through this interface it asks the QoS manager to check the wireless resources. After converting the RSVP request to a QoS class, QoS manager requests WFMP to reserve the connection from the AP. This request (identified by flow type 5 inside WFMP) automatically triggers flow detection in the flow classifier.

The flow identifier is given by the RSVP in filter spec and session parameters. The filter spec consists of the source IP address and source port/flow label values. Session contains destination IP address, protocol id and destination port values. The information carried by the filter spec and the session parameters is given to the WFMP so that it can identify actual data packets belonging to the RSVP reserved flow. Using this information WFMP can route the data packet belonging to a specific RSVP flow to the correct RAN connection. It is assumed that the RSVP – mobile IP addressing problem is solved. The problem appears if a mobile terminal in a visiting network is

identified by the care-of address and RSVP uses the home address. In such a case, the flow identifier given by the RSVP does not match to the identifier carried by the data packets. Thus, the packets belonging to the RSVP reserved flow can be neither identified nor routed to the correct radio flow.

Like the WFMP detected flow, RSVP flows are monitored by the flow classifier. Monitoring is needed to detect when the flow shall be released. Another option is that RSVP explicitly releases the flow by sending a teardown message to the QoS manager module. QoS manager then informs the WFMP to release the flow from the AP. RFC2205 says the following: "Although it is not necessary to explicitly tear down an old reservation, we recommend that all end hosts send a teardown request as soon as an application finishes." Due to this uncertainty, it must be possible to release the flow either through RSVP signalling or due to time out (after expiry of an inactivity timer).

Handling of differentiated services packets

Handling of the differentiated services (DiffServ) packets is another special case. WFMP must be able to detect DiffServ packets and treat them according to the value carried in the DS field. Each DiffServ packet is handled separately and, in addition, flow detection/monitoring is performed for DiffServ packets to detect flows.

When WFMP receives a DiffServ packet that does not belong to a flow, WFMP reads the value of the DS field and selects the correct default flow (BE/medium/high priority). WFMP has the knowledge how to map between the DS field values and the default flows. This knowledge is configured statically in WFMP to minimise interactions with the QoS manager module, i.e. WFMP does not need to consult QoS module each time a DiffServ packet is received.

Flow management at the M-Router 52

In this preferred embodiment, the M-Router is responsible for the flow detection and management of RAN flow identifiers. It also has an interface to the radio access network through which it can create and remove radio connections, and to QoS manager to retrieve QoS class and bandwidth estimates for new flows

WFMP uses two tables for flow management: active flows table (AFT) and default flows table (DFT). The detailed structure of the tables is presented in Figure 17. AFT is used to manage all the detected flows, whereas the DFT contains an entry for each registered mobile terminal. Default flows table makes it possible to route non-flow and DiffServ packets to the mobile terminals.

The active flows table has an entry for each detected flow. Depending on the flow type, the correct IP header values are stored in the AFT (for flow types and the corresponding parameters, see Chapter 0). In addition to the IP header values, the RAN flow identifier (RAN_ID) and the AP interface values are stored in the AFT. The idea is that once the flow type and the corresponding IP header values match, WFMP reads the RAN_ID and AP_if values from the table and forwards the packet to the correct RAN flow.

The default flows table is much simpler, it just contains the mobile terminal identifier (the mobile terminal IP address) and the RAN_ID and AP interface values. There is one RAN_ID value for each default flow; RAN_ID_1 for the BE traffic, RAN_ID_2 for the medium priority traffic and RAN_ID_3 for the high priority traffic. The values are selected by the WFMP during the mobile registration. If no entry from the AFT is found for an incoming IP packet, the packet is compared against the DFT. Once either the destination or the source IP address match to the mobile terminal identifier, WFMP reads the correct RAN_ID (RAN_ID_1 being the default choice) and AP interface values

from the DFT and forwards the packet to the correct radio flow. In case the non-flow packet requires a specific handling, RAN_ID_2 or RAN_ID_3 is selected.

Flow management in the downlink

First, normal IP routing methods are applied and the incoming IP packet is routed to the correct IP application/interface. The following rules should be followed when routing the IP packets:

1. If the packet is addressed to the M-Router itself (e.g. ping), it is processed in a normal way
2. Separate RSVP control packets from the other IP packets (identified by the RSVP protocol number 46)
3. Packets addressed to mobile terminals are sent to the WFMP process.

Next, a detailed description of the WFMP operation is given.

WFMP needs to select the correct RAN connection for incoming IP packets. If the packet does not belong to a flow, WFMP just selects the correct best-effort channel and sends the packet to the SAR. In case the packet belongs to a detected flow, the packet is passed to the compression module which then sends the packet to the correct RAN connection. Figure 18 clarifies the WFMP routing.

First, WFMP checks the flow information the incoming packet is carrying. It goes through the packet, including the extension headers, and saves all the relevant information. From the basic IPv6 header, source and destination IP addresses together with the flow label and traffic class (DS field) are saved. If the IP packet carries an ESP extension header (Encapsulating Security Payload), meaning IP encryption is used, the security index parameter (SPI) is saved. In case the packet does not carry the ESP extension header, and

TCP/UDP header is found, WFMP saves the port information and the protocol identifier.

After saving the flow information, flow classifier is called for flow detection purposes. Flow classifier detects one of the following cases:

1. The flow has already been detected, so just the flow information is updated
2. No flow has been detected yet and just the flow information is updated
3. Flow detection algorithm decides to create a new flow, starting from this particular packet.

When a flow is detected WFMP allocates a new RAN flow identifier for the flow. Since no explicit QoS or traffic parameters are available, WFMP has to communicate with the QoS manager to get the missing information. When requesting the QoS class and bandwidth estimate, WFMP gives the flow identifier (information used for flow detection) and TCP/UDP port information (if available) together with some measured traffic characteristics to the QoS manager module. Using this information, QoS manager calculates the QoS class and bandwidth estimation for this flow.

After the QoS manager returns the QoS class and the estimated bandwidth, WFMP reserves the connection from the AP and informs the mobile terminal and the compression module of the new flow. Finally, WFMP updates the active flows table (AFT).

Next, WFMP starts routing the packet to the correct radio connection. In case the flow has been detected, the correct RAN_ID can be read from the active flows table (AFT). The correct RAN_ID is found by comparing the flow information (according to the flow type) and the flow type to the corresponding values in the AFT. When all these fields match, WFMP reads the RAN_ID and the M-Router interface from the AFT and forwards the packet to the

compression module. When passing the packet to the compression module, also the RAN_ID must be given. After compressing the packet, the compression module sends the packet to the correct RAN connection (given by the WFMP).

If no entry from the AFT is found, the packet does not belong to a flow and is sent on one of the default channel. The correct default channel is found from the default flows table by comparing the destination IP address of the incoming packet to the mobile terminal_id values in the DFT. Each mobile terminal registered to the network has an entry in the DFT. Once the ids match (dst_addr = MT_id), flow classifier reads the corresponding RAN_ID (RAN_ID_1 if no special treatment required) and M-Router interface values from the DFT and sends the packet to the correct default channel. If the destination id does not match to any of the MT_id values in the DFT, the mobile is not registered to the network and the packet is discarded.

Flow management in uplink

When a packet is received from the radio access network, it is addressed either to a fixed host, or to another mobile terminal.

According to the RAN_ID, the SAR layer passes the received IP packet to the correct module. This is done automatically since there is a one-to-one relationship between each RAN connection and one of the modules (identified by the SAR SAP). The binding is done when the RAN connections are created. If the packet belongs to a flow and the destination is in the core network, the packet is passed to the compression module for decompression purposes (see Figure 19).In case the packet is non-flow traffic it is passed directly to the WFMP process. If the packet is RSVP signalling it is sent for RSVP module, and so on.

The uplink WFMP processing is somewhat different from the downlink operation; WFMP receives only non-flow packets and 'flow' packets

addressed to a fixed host. Therefore, the WFMP routing is much simpler compared to the downlink case.

The beginning is, however, similar to downlink operation, i.e. the flow information is saved and the packet is passed to flow classifier for flow detection purposes. The same three cases apply here: (1) flow already been detected, (2) flow not detected yet, or (3) new flow detected. If a new flow is detected, WFMP needs to select the new RAN_ID and request the QoS class and bandwidth estimation from the QoS manager. Then, WFMP informs the compression module of the new id. Also the AP and the mobile terminal must be informed of the new id, and a new service access point (SAP) added to the SAR interface. After the mobile terminal receives the flow information, it starts to use the new RAN_ID for all the packets belonging to that flow.

Option (1) always means that the correspondent node is located at the IP network. This is because mobile to mobile flows are switched directly between the AP interfaces. Options (2) and (3), on the other hand, require specific attention.

In both cases, WFMP has to detect mobile to mobile calls and forward the packet to the correct default channel. The correct default channel is found out by comparing the destination IP address to the values in the default flows table. In option (3), the packet is first sent on the default channel (best-effort) and only after that the flow is switched. If the destination is located at the IP network (not found from the DFT), the packet is sent to the IP forwarding process which forwards the packet to the correct network interface (normal IP routing applied).

Mobile to mobile traffic

The start is similar to uplink operation, that is the SAR first forwards the packet to the correct process depending on the RAN_ID.

In case of non-flow traffic, the SAR sends the packet to WFMP. Since the packet is addressed to another mobile terminal, WFMP finds an entry from the default flows table. Then it simply forwards the packet to the correct default channel (read from the DFT). MT-to-MT non-flow packets are always routed via the WFMP.

When the WFMP detects a new flow and the destination is another mobile terminal, WFMP creates a connection directly between the two mobile terminals. This means that WFMP selects new RAN_IDs for both AP links (one for MT1, another for MT2), informs the receiving and sending MTs of the new ids and adds the new connections to both APs. The compression module is not informed since the compression is used between the MTs directly. Mobile to mobile routing case is clarified in Figure 20.

First, the receiving mobile terminal is informed of the new RAN_ID. It then sends the acknowledgement back, meaning that the receiver is listening the new RAN_ID. Then, the new RAN_ID is given to the sending mobile terminal. The sending mobile terminal understands that the packets belonging to that flow must be compressed. It just starts using the new RAN_ID and applies the compression mechanism. Finally, the receiving mobile terminal receives the first packet carrying the full IP header. It must then save the header since the consecutive packets are compressed.

Since the detected MT-to-MT flows are not transmitted through the M-Router WFMP module, the WFMP cannot monitor the active flows. This means that it cannot detect when the flow should be released. Therefore, the receiving mobile terminal side has to control the traffic and inform the M-Router WFMP when a flow should be released. Possible monitoring places are the MT WFMP or MT FC modules. An optional solution is to monitor the traffic in the SAR block. The SAR block could measure the traffic going through the specific RAN connections (identified by the RAN_ID) and after a constant time of inactivity (time out) SAR would inform the WFMP to release the flow.

Flow management at the mobile terminal

Flow management at the mobile terminal (MT) side is much simpler than in the M-Router side. The MT WFMP does not detect flows, it just starts using new radio connections when commanded by the M-Router WFMP.

Flow management in uplink

MT WFMP just needs to pass the IP packet to the correct RAN connection. Packets not belonging to a flow are sent on one of the default channels, packets belonging to flows to the correct radio channel according to the AFT (see Figure 21) Like at the M-Router side, WFMP first updates the flow information (the same four flow types identified) and uses the correct parameters in reading the AFT. If the flow information and the flow type correspond to one of the AFT entries, the packet belongs to a flow. Otherwise, the correct default flow is read from the default flows table.

Flow management in downlink

The downlink case is also quite simple (See Figure 22). WFMP does not need to do any routing between RAN channels, it just passes the received IP packet to the IP forwarding which finally delivers the packet to the correct application. Like at the M-Router side, the SAR already takes care of the routing by passing the incoming packet to the correct MT process.

Since the M-Router WFMP cannot monitor mobile to mobile flows (flows are switched directly between the M-Router AP interfaces, thus bypassing the M-Router WFMP), the MT WFMP has to monitor incoming flows. This means that when the MT WFMP notices that the flow does not exist anymore, it has to inform the M-Router WFMP to release the flow from the RAN.

There are several different mechanisms that could be applied in the WFMP flow detection to that described above. For example, the following ideas can be used if minimal processing load is required in the M-Router.

The simplest mechanism is based on IPv6 flow labels and X/Y classifier. WFMP just checks the flow label value, and if it is non-zero uses the source IP address and the flow label for flow detection. In case the flow label is zero, the packet is treated as best effort traffic and no flow classifier is applied (= no flows detected for packets having zero flow label).

Flow compression Scheme

Header compression may be performed for IPv6 as described in the Internet draft on IP header compression by Degermark. The method for grouping IP datagrams into compressible streams discussed in the document will not be implemented. It is assumed that the WFMP module in the M router performs the grouping of the IP datagrams on behalf of the flow compression (FC) module.

Every terminal that enters the network is assigned a best effort (BE) channel. The datagrams sent on the BE channel are not compressed due to lack of similarity between individual datagrams. Whenever the M router's WFMP module identifies a new flow it sets up a new radio channel for that flow and the datagrams sent over the channel are compressed by the FC.

Whenever an IP datagram is sent either uplink or downlik, a check is made to see if it belongs to some flow. If the flow exists, the IP datagram is sent to the FC module, which performs the header compression and sends the compressed IP datagram via the assigned radio channel to its destination. If the datagram doesn't belong to any flow, it is sent uncompressed on the BE channel. Whenever a new flow is detected, a new radio channel is allocated and the FC is notified to set up the compression state and to direct the incoming datagrams from the radio flow to the FC. When the flow ends, the FC is notified so that it may deallocate memory reserved for bookkeeping purposes.

The header decompression is somewhat simpler than the compression. Whenever data is received from a BE channel, it is delivered directly to the WFMP. If the radio channel belongs to a flow, the data received is sent to the FC which uncompresses and forwards the IP datagram to the WFMP. When the flow ends, the FC is notified so that it may deallocate memory reserved for bookkeeping purposes.

Even if header compression is done on every non-BE channel it is possible to use direct point-to-point flows ("cut-through") between two terminals without going through the M router and its FC.

According to the draft proposed in Degermark, the packets are partitioned into two different categories, one for TCP and the other for non-TCP packets. The packets in each category are further divided into packet streams based on IP addresses, port numbers, etc., a task that is already performed by the WFMP software module in the M-Router. This partitioning between TCP and non-TCP packets together with a context identifier (CID) uniquely identifies which compressed and uncompressed packets belong to the same packet stream. The compression is done by sending only those header fields that change, or in the case of TCP by sending the change from the previous datagram. The draft examines which header fields may be inferred (e.g. packet length), which are constant, and which header fields have to be either sent as-is or as a difference from the previous.

Four new packet types are defined in addition to the normal IPv6(/v4) packets:

The full header packet indicates an uncompressed packet that belongs to a compressible stream. It includes the context identifier (CID) and a generation for non-TCP packets coded into the length fields present in the packet header.

The compressed non-TCP packet includes the CID together with the generation and the fields that have changed since the previous full header packet, which is identified by the generation value.

The third type is the compressed TCP packet that "...indicates a packet with a compressed TCP header, containing a CID, a flag octet identifying what fields have changed, and the changed fields encoded as the difference from the previous value", i.e. the previous packet. The TCP checksum is also included in the packet.

The fourth type is the compressed TCP nodelta packet that is similar to the compressed TCP packet except that the header fields sent as the difference from the previous packet are sent as-is. This type of packet is only sent in response to a header request issued by the receiver.

These new packet types are indicated by sending a specific type value either on the link layer level or by adding an additional byte in front of the compressed packet. It is also assumed that the length of the packet is given by the link layer.

The compression is started by choosing a suitable CID value and sending a full header to the decompressor. Full headers are sent with an exponentially increasing period until an upper time or packet limit is reached. To recover as quickly as possible from TCP packet errors the decompressor may request the full headers for a set of TCP CIDs. The request is sent as a context state packet with a list of CIDs which won't decompress correctly which means that synchronization is lost between the compressor and the decompressor. The compressor replies with a new TCP nodelta packet for each CID requested.

Qos management

QoS is a new trend in IP-networks. Formerly QoS has been realized by ATM, but increasing amount of IP applications (users) demanding QoS from the network, has forced network designers to pay attention to QoS in IP networks.

The present system is designed to take advantage of customer -and core network's QoS mechanisms. Typically, Integrated services based mechanisms are seen as QoS mechanisms in the last hop of the network. Differentiated services based mechanisms are seen more as core network mechanisms. Both mechanisms have been considered in developing the present system. Also some other mechanisms to separate packets from others have been introduced.

General QoS management concept

In practice, QoS means differentiating classes of data service - offering network resources to higher-precedence service classes at the expense of lower precedence classes. QoS also means attempting to match the allocation of network resources to the characteristics of specific data flows [QoS]. These ideas are deployed in the present system.

QoS can be implemented by differentiating data flows in the basis of different information: IPv6 Flow-ID + source address + destination address, port information + source address + destination address, Priority bits + source address + destination address, RSVP reservations or H.323.

These flows can be treated differently from each others, and QoS can be implemented by multiplexing these flows in the basis of QoS parameters of each flow. These parameters can be explicit values (peak cell rate, bandwidth requirement etc.) or simply an information of preferred Class of Service. It depends on the mechanism how the QoS parameters are determined.

Packets belonging to a flow, are put to a proper radio queue. In the preferred embodiment of the present invention, there are three different queues: Best Effort, Controlled Load and Guaranteed Service. These flows get different priority from each others, and also scheduling inside the queues will be performed.

QoS Manager (QoS Entity)

QoS manager's main task is to map fixed network's QoS parameters to radio QoS and communicate with radio resource manager. In practice this means mapping explicit QoS values to radio priority queues. QoS manager has to know some statistics of flow, and proportion this to available radio bandwidth. With this information QoS manager can prioritize different flows.

QoS Manager has interfaces to RSVP, H.323 and WFMP entities (as explained above with reference to Figure 15). These interfaces and main signals are presented in following table:

Table 4: QoS manager interfaces

Interface	Signals
M-ROUTER_QoS –>M-ROUTER_RSVP	RESV_FLOW_conf RESV_FLOW_req
MT_QoS ->H.323 H.323 –>MT_QoS	SETUP_conf, CLOSE_conf, UPDATE_conf SETUP_req, CLOSE_req, UPDATE_req
M-ROUTER_QOS –> M-ROUTER_WFMP M-ROUTER_WFMP –> M-ROUTER_QoS	RESERVE_FLOW_req RESERVE_FLOW_conf RR_STATUS_enquiry
M-ROUTER_QOS –> M-ROUTER_WFMP	UPDATE_req UPDATE_conf

M-Router's QoS manager has more functionality than the mobile terminal's QoS manager, because flow establishing is performed in M-Router. The most important functionalities of the mobile terminal's QoS manager is to assist H.323 signalling and handovers.

Active flows table

Active flows table is the place where all information concerning a data flow is stored. This table is accessible for WFMP and QoS Manager. As mentioned above, an active flows table is exemplified in Figure 17. In addition to proposed Active Flows Table, QoS information also needs to be included in that table, together with an indication whether the flow is signalled or only detected by WFMP.

Ways to get QoS information for a connection

RSVP (Figure 24)

RSVP is a resource reservation protocol, which tries to reserve bandwidth and desired QoS for a particular data flow. This system is readily supported in the present system, because flows are detected by WFMP, and flows can be treated differently from each others in radio link. RSVP uses control messages for marking reservations in intermediate network elements. These control messages are separate from application data. These messages are separated from other data in the basis of protocol number, and directed to RSVP entity before WFMP process.

Two different scenarios exist when using RSVP:

1. WFMP has already detected a flow, and created a dedicated channel for flow and after that RSVP entity gets reservation request for that particular flow.

2. RSVP entity gets reservation request before WFMP detects the flow. In this latter case RSVP should trigger WFMP. This can be done via QoS manager.

Figure 24 shows how RSVP reservations are handled in the present system. This particular picture presents situation where MT is receiver and sender is somewhere in network (downlink case).

RSVP messages (PATH/RESV) use protocol number 46, and that is how reservation messages can be separated from best-effort traffic. These messages will be delivered to RSVP entity which handles them. RSVP entity talks with QoS manager which asks WFMP to establish flow with appropriate QoS values. WFMP asks resources from RRM, and QoS Manager doesn't have worry about this.

In Mobile IP Router, RSVP Entity has two roles, it acts like a normal RSVP Daemon, but also makes wireless specific operations. Normal RSVP Daemon checks the capacity of the Mobile Router itself, and forwards / manipulates RSVP-messages in IP level. Wireless RSVP Daemon communicates with WFMP and ask it to establish flows with certain parameters.

RSVP RESV messages can also be so called refresh messages, which are sent periodically. These messages shouldn't trigger a new flow, but only refresh the existing. This is done in M-ROUTER-WFMP in following way:

1. RESV refresh message triggers QoS manager to send RESERVE_FLOW_req to WFMP.

2. WFMP checks from active flows table if it already has signalled flow for that particular dataflow. In the Active Flows table, there has to be an indication if the flow is signalled by RSVP or detected in some other basis (like traffic volume).

3 If flow already exists, WFMP only sends confirmation message, and makes no other actions.

Differentiated Services

Differentiated services means generally deployment of priority bits in every IP-header. If WFMP detects a flow of IP-packets with priority, it should inform QoS Manager about these bits. QoS Manager includes functionality that understands the bits, and makes a mapping to required Radio QoS. Pameters are marked to the Connection table for that specific flow.

How parameters are mapped into explicit QoS requirements, depends on deployed differentiated services in network side. When the formula of priority bits is ready, it's relevant to map the parameters into explicit Priority Classes. The standardisation of priority bits is still going, and there may be different ways to deploy priority bits in the future. Tables 5 and 6 exemplify how bits are mapped into priority classes in the preferred embodiment of the invention.

Table 5 Example of bit pattern

Bits	Indication
Bits 0-2	000 = Drop Preference 1, 001 = DP 2, ... ,111 = DP 8
Bit 3	0 = Normal Delay, 1 = Low Delay
Bit 4	0 = Normal Throughput, 1 = High Throughput
Bit 5	0 = Normal Reliability, 1 = High Reliability
Bits 6-7	Reserved for Future Use

Table 6 Example of mapping Priority bits to QoS Classes

Priority bits	QoS Class (see Table 8)
???001??	Class 3, BE
???101??	Class 2, Controlled Load
???110??	Class 1, Guaranteed Load

If WFMP detects packets that include priority bits, but can't detect a flow from that traffic stream (not enough packets per second), WFMP puts these packets to the right priority queues. In other words, WFMP does not ask

anything from QoS manager, if packets are only occasional. That is, the intelligence is split between WFMP and QoS-manager, so that the QoS – manager does not become over deployed.

Well known ports

There are many "well known" TCP /UDP ports, indicating that traffic needs some real time features, or it may also indicate that the amount of traffic is extra high or low. This kind of ports are e.g. ftp-port or telnet port, which both have very different characteristics. Ftp needs much bandwidth, but it isn't so critical with real time requirements. On the opposite, telnet doesn't need much bandwidth, but it shouldn't get affected by high delay. This information can be deployed when choosing the right radio link queue for the data flow. In Preferred system, it's relevant to take advantage of port information after a flow has already been detected. This means that port information itself doesn't trigger WFMP to notice a new flow, but after a flow has been detected, port information can be deployed. If IPSEC or some other protocol hides port information, then it can't be deployed.

Figure 26 presents a situation, where WFMP has detected a flow and port number belongs to an applications that is identified by QoS Manager.

Table 7 includes some common ports that could be treated differently. Listing of ports is only an example, and the ports that get specified service, could be changed.

Table 7 Ports (Example)

Type	Port number	Explanation	Possible QoS Classes (See Table 8)
ftp-data	20/tcp	File Transfer [Default Data]	Class 3
ftp-control	21/tcp	File Transfer [Control]	Class 3
telnet	23/tcp	Telnet	Class 2

http	80/tcp	World Wide Web HTTP	Class 2
snmp	161/tcp	SNMP	Class 3
ipx	213/tcp	IPX	Class 3
dhcpv6-client	546/tcp	DHCPv6 Client	Class 2
dhcpv6	547/tcp	DHCPv6 Server	Class 2
vat	3456/tcp	VAT default data	Class 1
Vat-control	3457/tcp	VAT default control	Class 1

The network administrator is preferably able to configure ports that get special handling as it is desirable to configure the classification in the basis of what customer company needs. Some company may use multimedia applications much more aggressively than others. Also, some companies may use their own specific applications that should get most of the bandwidth (e.g. banks). This special treatment is possible, for example, if the QoS Manager is a separate functional entity that can be updated easily.

H.323

In the system of the present invention H.323 can be deployed in two different ways: 1) H.323 applications signal connections via the QoS manager, or 2) H.323 signals connections by using RSVP (Figure 27). In both cases, there also has to be a mechanism to update the connection.

The H.323 call signalling procedures are made of five steps from call set-up to call termination. The call set-up procedures with all possible cases are complex and are only briefly outlined below.

Firstly, a SETUP message is sent from the calling endpoint to the other party which responds with a CONNECT message. Next, both parties exchange system capabilities by transmission of the H.245 TERMINAL-CAPABILITY-

SET message. During the third phase logical channels for the various information streams are opened using H.245. These streams are transported over an unreliable protocol specified by H.225.0. Data communications which is transmitted in the logical channels set-up in H.245, are transported using a reliable protocol (H.225.0). During a session, the procedures for changing capability, receive mode etc. are specified in H.245.

The bandwidth required for the connection can be determined from the capabilities that are agreed to be used between the terminals with a table that maps the audio codec selected to the bit rate required by the codec. For video codec, the maximum allowed bit rate is included in the capability set message per available video codec.

Finally, the call is terminated by either of the parties with an END-SESSION-COMMAND in the H.245 control channel. The other party responds with a RELEASE-COMPLETE message.

The interface from the H.323 protocol stack to the QoS manager requires three functions to facilitate the above mentioned functionality with H.323. The required bandwidth with information about the parties (source & destination address) must be passed to the QoS manager after the call set-up phase. If the required bandwidth between the parties is changed during a call it must be signalled to the QoS manager as well with an update request stating the new bandwidth and identification information for this connection so that the QoS manager can identify which connection's bandwidth it has to modify. Finally, after the call is terminated the bandwidth must be released by signalling the QoS manager that this connection has been terminated.

RAN QoS functions.

The QoS based radio access network has to be able to provide bandwidth on demand, class based queuing and reliability. In a wireless transmission link multiplexing of different services into the medium typically requires consideration on four QoS accounts: bandwidth, delay, jitter, and reliability.

Bandwidth is the first requirement for QoS driven services i.e. to be able to support the requested traffic parameters. In the wireless link the main objectives are efficient channel utilisation while maintaining service specific QoS for TCP/IP traffic. This means that the AP Scheduler should know the requested average and/or peak bandwidth of those connections for which the radio flow is to be established. This way the Scheduler can guarantee the satisfaction of bandwidth on demand and perform statistical multiplexing.

Delay and *Jitter* are primarily affected by the traffic scheduling over the wireless link. In the present approach the flow based connections are queued separately (queue for each connection) and connections are grouped into 3 different delay class queues. In order to be able to put the packets to the right queue, the Scheduler (or queuing function) needs to know the flow ID and delay class of the incoming packet. Also, to take the delay and jitter requirements into account in choosing the packets to be sent, the Scheduler should know a) maximum allowed delay of the packets at RAN layer b) keep a time stamp for each packet.

Reliability over wireless link requires error control which is typically given for instance by coding and/or data re-transmissions. Coding is used both for the error detection and correction which imposes constant overhead over the applied data. ARQ (Automatic Retransmission reQuest) is only applied for the corrupted packets which is feasible as long as the packet loss probability is not too high and delay of retransmission is admissible. The scheduler needs

information about ARQ usage per radio flow (connection) basis. (e.g. No ARQ, Limited ARQ, ARQ). FEC usage can be fixed, used for all packets.

Table 8 presents an example of mapping from TCP/IP QoS into the radio access network specific QoS according to a preferred embodiment of the present invention. The first two columns specify radio access queuing and error control while columns 3-5 show different TCP/IP level QoS concepts.

Table 8. Example of network QoS mapping into radio access QoS.

Delay Class	*Radio Access QoS*	*Transmission Protocol*	*Integrated Services*	*Differentiated Services*
1st class	*No ARQ+FEC*	-	*Guaranteed Load*	*low delay/high dropping*
2nd class	*Limited ARQ+FEC*	*UDP/RTP flow*	*Controlled Load*	*medium delay/ medium dropping*
3rd class	*ARQ+FEC*	*TCP flow/No Flow*	*Best Effort*	*high delay/low dropping*

In the preferred embodiment, the M-Router functions as a central intelligence point of the radio access network detecting flows, classifying them and mapping network QoS concepts into radio QoS capabilities.

Queuing (See Figure 28): Queuing strategy for priority classes 1 and 2, in the preferred embodiment, is based on the radio flows such that each radio flow has its own queue. Based on the flow ID, the right priority class can be chosen as well as the queue where the packet is put. This approach is required because the Scheduler has to be able to differentiate the

connections and their QoS requirements. For Best Effort data (priority class 3), the flows may also be identified.

Delay and Jitter: These are primarily affected by the error protection scheme and traffic scheduling over the wireless link. It has been found that adding one more queue improves the service quality that can be provided for Internet voice, thus increasing the service differentiation capability.

Scheduler

The wireless environment puts a special stress also on the performance of the scheduling algorithm. This requires a scheduling algorithm that is efficient and aware of QoS and traffic characteristics of the connections..

The scheduling algorithm has an important role in controlling the flow of the packets over the band-limited wireless channel. Used together with Call Admission Control (CAC) and resource allocation, the scheduling can be used to guarantee the satisfaction of different QoS requirements for different traffic types. Admission Control and resource allocation operate at the time of the connection is established, deciding whether new radio flows/connections can access to the channel. The scheduling makes the decision on choosing the packets to each MAC frame. The scheduling algorithm should aim to provide the following properties: [Garrett 96]

- Maintenance of traffic characteristics of the connections
- QoS requirements satisfaction - the QoS parameters related to delay and loss are important to maintain according to the traffic contract.
- Statistical multiplexing gain - the scheduling should smooth or take into account the effect the connections with variable bit rate have on the buffer occupancy (congestion).

- Utilization of bandwidth unallocated or allocated to idle connections - since the applications (WWW-browser for instance) may be not sending cells all the time, being silent, the unallocated resources should be utilized during these periods.

- Declared and real traffic consistency - in the case where the source is producing more traffic than it's expected and thus breaking the traffic contract, the scheduler should for instance 'drop' the priority of the connection.

With the queuing scheme of the preferred embodiment, the Scheduler could work for instance in the following way:

Firstly, the Scheduler prioritizes the packets according to the three priority classes. Class 1 has highest priority and class 3 has the lowest priority. Scheduler begins to allocate packets pending in the class 1 queues. Inside the priority class, the prioritising between packets/flows can be made according to delay requirements i.e choosing packet who have least 'lifetime' left and so forth. Parallel to this the flows consuming less bandwidth than allocated have higher priority. This can be taken into account by using traffic policing function such as Token Bucket. When all the packets of the priority class 1 have been allocated and there's still room in the MAC frame, the Scheduler begins to allocate packets from priority class 2 queues in the same way as in the class 1 case. After all the class 2 packets have been allocated, the Scheduler allocates class 3 packets into the free slots. Class 3 queue works as a FIFO (First In, First Out). The scheduling ends when all the packets have been allocated or when the MAC Time Frame is full.

Mobility management

Terminal Registration and Authentication

The terminal performs the IP level registration process when it has been powered on. It is also the initial part of an IP level handover. The process is the same in the terminal's home network and in the foreign networks.

The process is performed after the link level registration process. The link level procedures have already authenticated the terminal and accepted its access to the network. The link level entities have also made the terminal a member of the all-nodes and solicited-node multicast groups. The solicited-node multicast address is calculated from the EUI-64 identifier provided with the link level registration messages.

These multicast groups are local to the link that the terminal is currently attached to. This means that both the terminal and the M-Router know that this terminal belongs to these multicast groups. The M-Router knows to route packets that are addressed to these groups to this terminal and the terminal is able to receive and process these packets. This is why the terminal does not perform an explicit join to these groups using the link level multicast membership protocol in the beginning of the IP level registration process after it has generated its link-local IP address.

When the link level has performed its registration procedures the link level informs the upper level entities of its readiness. If this is a power on situation the network interface in the terminal becomes into an enabled state. In this case the terminal generates its link-local IP address from the information provided by the link level (an EUI-64 formatted MAC identifier of the interface). Normally a host would validate this address by performing duplicate address detection procedures before the final assignment of the address. However, the MAC identifier has already been verified during the link level registration procedures so there is no need to carry out this verification

again as the link-local IP address is generated from that same unique identifier. The terminal assigns the link-local IP address to the network interface.

The following actions are performed in both power on situation and in case of IP level hand over. The host would normally use link level mechanisms to join now to the all-nodes and the solicited-node multicast groups. For the reasons mentioned above this task is not performed in this environment at this point since the task has been carried out implicitly during the link level registration procedure.

Once the network interface has been assigned a valid link-local IP address the terminal performs Router Discovery actions in order to find its default router and possibly to get the network prefixes for its site-local and global IP addresses. The M-Router answers to the terminal's solicitation by an advertisement and provides the terminal with information on M-Routers link level address as well as its IP address. The terminal updates its default route to point to this address. Note that the Router Discovery process is necessary even in the home network because the home network prefixes might have been renumbered.

The router may advertise the site-local and global IP address prefixes for this link in the advertisement. If this is the case the terminal generates its site-local and global IP addresses based on these prefixes. Again, the terminal does not need to verify the uniqueness of the addresses by duplicate address detection procedures since the network interface specific suffix of the addresses has already been proven to be unique during the link level registration process. Also, in this case the terminal will probably not need to join to the solicited-node groups of these addresses as these groups are potentially the same as the solicited-node group for the link-local address of this network interface. The terminal assigns its site-local and global IP addresses to the interface.

The bits in the Router Advertisement may also be set so that the terminal is required to acquire its site-local and global IP addresses by stateful configuration mechanism such as DHCPv6. In that case the terminal sends a solicitation to the all configuration servers group on the link in order to find a server that is willing to serve the terminal. In this environment the server would reside in the M-Router. The server responses with an advertisement so that the terminal would know the IP address of the server. The terminal sends a configuration request to this address and receives the requested addresses in the servers reply. The configuration server would probably add the terminal into the solicited-node groups of these addresses on behalf of the terminal. Otherwise the terminal would have to initiate the generic link level and IP level multicast group membership procedures in order to join to these groups. In any case the terminal must enable the receiving of the packets coming from the interface and addressed to these solicited-node groups. The addresses need not to be verified by the duplicate address detection procedures, the configuration server and the other entities within the M-Router are supposed to know which addresses are valid for the terminal's interface. The terminal assigns its site-local and global IP addresses to the interface.

The stateful configuration mechanism can be optimized further by leaving out the DHCP Solicit and the DHCP Advertise messages. They are not needed if we assume that the stateful configuration server is located in the M-Router, whose address we know already from the Router Discovery process. This would be a non-standard deviation from the stateful configuration mechanisms and may not go in hand with the proposed IP level authentication methods.

Addressing Scheme

IP level addresses are constructed based on the standard address configuration methods. The network interface specific part of the IP addresses is formed from the link level MAC identifier of that interface. The identifier is in

the EUI-64 format. This identifier is verified during the link level registration process and is proven to be unique among all the terminals that are attached to the same subnet. If the verification fails the terminal is not able to use the network.

The terminal's specific network interface's link-local IP address is generated from a static well-known prefix and the network interface identifier. The link-local address is not verified by the Duplicate Address Detection process.

The network interface's site-local IP address is formed from the site-local prefix received in the Router Advertisement and from the network interface identifier. The site-local prefix uniquely identifies this subnet within the site area. The site-local address may also be acquired from the stateful configuration server. In both cases the generated or received address is assumed to be valid and no Duplicate Address Detection is performed.

The network interface's global IP address is either generated from the global prefix provided by the Router Discovery process and from the network interface identifier. The global address may also be received from the stateful configuration server. The global address is not verified by the Duplicate Address Detection process.

The information in the M-Router that is used in the terminals' address allocation process (either stateless or stateful) is managed by methods that are outside the scope of this document. It is also assumed that the terminals are attached to just one link at a time so that only one M-Router is accessible at a time. The addressing scheme deviates from the standard procedures by the fact that no Duplicate Address Detection is performed.

Location management (RAN + IP)

When the terminal has moved to another subnet the link level performs its registration procedures. The IP level is informed about this by the link level as

described above (Terminal Registration and Authentication) after which the IP level performs its own registration procedures as described in the same section. The IP level mobility deviates from the standard with the fact that an indication is received from the lower layer each time the subnetwork changes.

Although the potential subnetwork change has been indicated by the link level to the IP level, the ultimate fact that the terminal has moved to another subnet is deducted from the site-local and global address prefixes acquired from the Router Advertisements or from the site-local and global addresses received from the stateful configuration server. These new prefixes are compared to the old ones that were active on this network interface before the movement. If the prefixes differ the IP mobility procedures must be activated.

The terminal uses a special header extension in the terminal originated packets when the terminal is away from its home network. The header extension is called the Home Address option. The Home Address option contains the home address of the terminal's network interface. This option makes it possible for the terminal to use its new global address in the packets' source address field so that the packets can more easily penetrate the firewalls along the path. The Home Address option is handled by the receiving nodes so that the source address of the received packet is replaced with the address in the Home Address option. This way the ongoing sessions are not disturbed even when the active address of the terminal is changing.

The terminal informs its previous default router of its new global IP address by sending a Binding Update to the router. The message is authenticated with the IPSEC AH header. The previous router acknowledges the update. The previous router becomes a proxy for the terminal's previous global address. This means that the previous router captures all the packets that are destined to the terminal's old global address and encapsulates each of them into a new packet and sends it to the terminal's new global address. The previous router also disables the allocation of the terminal's old address for the other

terminals under its area until the Binding Update expires. When the Binding Update expires the previous router will no longer act as a proxy for the terminal's old global address.

The terminal also informs the router in its home network of the terminal's new global IP address. This is performed by the same Binding Update + AH/ Binding Acknowledge + AH transaction as before. The home router now acts just as the previous router in the previous case. The only difference is that the terminal sends Binding Updates regularly to the home router so that the Binding Update never expires.

The terminal informs also all the corresponding nodes it has recently been communicating with of its new location. Again it uses the Binding Update + AH message. The corresponding nodes update their data structures so that a Routing Header is added to each packet that is destined to the terminal's home address. The Routing Header contains the new global address of the terminal's interface. This makes the packet to be routed first to the new location of the terminal. There the Routing Header is removed and the packet appears as if it had arrived to the terminal interface's home address. As the home address is still a valid address of the terminal's interface, the packet can be received and processed as if the terminal was in its home network.

Inter IP subnetwork handover within router

This means the case when the terminal moves between router ports which belong to different IP domains. Here, mobile IP allocates new address but flows can be maintained in RAN. In the IP level this kind of handover is performed exactly the same as in the previous case.

DHCP v6

When a host wishes to acquire a global Ipv6 address and has received a router advertisement with the M bit set, the host has to follow the principles of stateful address autoconfiguration, i.e. DHCPv6.

The stateful address autoconfiguration starts with the host, i.e. the client, sending a DHCP solicit message to the all DHCP agents multicast address in order to find one of the site's DHCP servers. The DHCP server replies with a unicast DHCP advertise message that contains the server's IP address. Thereafter the client sends a DHCP request to the server in order to obtain the network's parameters. The server responds with a DHCP reply wherein a variable amount of parameters are delivered to the client. The client may end the DHCP session by sending a DHCP release message to the server which then acknowledges the release by sending a DHCP reply. The server may also notify the client if some parameters change by sending an DHCP reconfigure message.

The advertise, request, reply, release and reconfigure DHCP messages are all unicast messages and may be followed by a variable amount of extensions that carry additional parameters between the client and the server. In addition to configuring the client's IP address and DNS entries the extensions include information about time zones, TCP parameters and other network information. The DHCP messages may also be authenticated.

The server and the client don't have to be on the same link, a DHCP relay server may be inserted between them to provide a larger domain served by only one DHCP server.

The DCHP server may attach an client key selection extension to its advertise message to indicate which security parameter index (SPI) value the client should use when authenticating itself to the server. The authentication is accomplished by adding a client-server authentication extension to any DHCP message. The purpose of the DHCP authentication extension is to provide

authentication in the case that the client does not have large enough address scope to reach the server from the beginning, i.e. a DHCP relay is used and therefore the normal IPSEC procedures can't be applied. The current draft assumes that the key used for authentication has to be known by both the client and the server before the DHCP registration procedure.

DHCP v6 in subnets

When DHCP is used to create the global IPv6 address it is very likely that the M router doesn't know the newly created global IPv6 address if the M router and DHCP server are separate entities. The DHCP server has given away an IPv6 address with the correct network prefix to the terminal which implies that the M router will sooner or later receive IPv6 datagrams destined to this address. In order to send the datagrams to the correct terminal via the correct VPI/VCI connection, the M router has to perform address resolution according to the neighbour discovery protocol [Narten98]. The M router simply sends a multicast neighbour solicitation on the link, and receives the link layer address in the neighbour advertisement sent in response by the terminal that has acquired the IPv6 address in question. The M router may thereafter send the datagram via the correct connection and add the global IPv6 address to its lookup tables.

Interoperation of RSVP with Mobile IPv6

In Mobile IPv6 a Mobile terminal (MT), which is visiting a foreign link, can be addressed either using its Home Address (HA) or using some of its care-of addresses (CoAs). Furthermore, if a Correspondent Node (CN) directs packets to the MT's HA (or to a stale CoA), the packets are tunnelled to the MT via the Home Agent (or a router at a previously visited link).

The various modes of addressing the MT raise issues of interoperability in a Mobile IPv6 environment, where RSVP ([RFC2205], [RFC2209]) is used for network resource reservation. The addressing modes must be taken into account when

- composing and transmitting RSVP messages,
- setting up the reservation state in the network nodes,
- classifying packets when implementing the traffic control, and
- consulting routing process for finding out outgoing interfaces and becoming notified of route changes

For conveying addressing information in RSVP messages and maintaining reservation state at the network nodes two options can be identified:

1) *The MT is always identified by its Home Address in RSVP messages*

This straightforward approach stems from the idea that a RSVP PATH message originated by the sender application (which only knows the MT by its HA) would not be changed on its way to the receiver. This means that RSVP state at all nodes would need to be enhanced to encompass at least two addresses for the MT: the HA and a CoA. This means that the nodes have to be informed of the address binding. The HA is needed for identifying the RSVP session when trying to find a matching path state for a RESV message. On the other hand, CoA would be used for packet classification (if any) and consultation with the routing process.

This approach would necessitate changes to all RSVP-aware routers. It would also violate the wording in the current specification, which requires that the destination IP address of a PATH message must be the same as the DestAddress of the session ([RFC2205] p. 36).

2) The MT is identified by its CoA in RSVP messages when the MT is not at home

As far as forwarding of application traffic is concerned, in Mobile IPv6 intermediate routers need not be aware of the binding between HA and CoA, except for the case of a Home Agent that is tunnelling packets to the MT. The same principle can be also applied to RSVP operation. This is facilitated by the observation that at the routers RSVP reservation state need not be maintained across the whole life span of an RSVP session. Instead, when the MT is roaming, a new path state can be built along the data path for each new CoA without any knowledge of the actual HA. This approach is more scaleable as it necessitates changes only at the end systems, while RSVP processing at the intermediate routers remains as specified in current RSVP RFCs.

To achieve smooth roaming, several message processing extensions and enhancements are required at the MT and the CN. These incorporate IP address mapping between HA and CoA with the support of IPv6 address configuration mechanisms and Binding Cache management.

Option 2) above would seem to be more feasible in operational networks having several core network routers. However, In the present system, any deviations from standard RSVP behaviour cannot be expected at the fixed network nodes. Therefore, option 1) has to be exploited with the assumption that the MT and the M-Router, which know the bindings between CoA and HA, are the only RSVP-aware nodes on the path between MT and CN.

Summary

The complete system described in the preferred embodiment is based on IPv6 network architecture in which time-critical IP flows are mapped into dedicated radio flows, and in which non-flow traffic is assigned to a default flow of appropriate priority. The mechanism described allows the user/applications to assign different IP QoS parameters for various flow types

over the radio link. The radio link QoS management described supports both integrated and differentiated IP QoS mechanisms.

However, as will be appreciated by a person skilled in the art, not all the components of this system are essential to the invention: some features are not required at all, whilst others are only exemplary and thus may be modified.

The present invention resides in the provision of a system which supports differentiated services. That is, a system which has a plurality of default radio flows with different (fixed) radio QoS characteristics. The solution described allows the operator and/or the manufacturer of the network to define the mapping between the fixed IP (DS fields) and radio QoS according to the desired flow policy of the particular network. For example, he can dynamically configure the flow class criteria (e.g. how many similar packets will be needed for detecting a flow) and assign his own policy which defines how the differentiation bits are mapped into the default queues.

In the preferred embodiment, the system has three priority queues, high priority service queue for premium class traffic, medium priority service queue for assured class traffic and low priority service queue for best effort class traffic, and each priority queue has preset radio scheduling parameters which define these QoS. However, it will be appreciated that the number of default channels depends on the network requirements, and may alternatively be more or less than three.

As indicated above, differentiated services generally means deployment of priority bits in every IP header. In the preferred embodiment, Tables 1 and 2 illustrate priority class bit mapping which could be employed in an IPv6 header. The standardisation of priority bits has not yet been effected, and it

would be clear to a person skilled in the art how this concept could be adapted to be used in a different header (for example the TOS octet of the header defined in IPv4 or one defined in a future standard).

Also, in the preferred embodiment, the WFMP entity (M-router 52/ mobile terminal 4) process the type of service field and categorise the different packets into one of the default radio queues and mark the IP packet with the corresponding RAN identifier. Then, the mobile terminal/access point decodes the RAN identifier and maps the packet into the corresponding radio flow. However, performance of these functions is not restricted to these components. For example, the mapping could be performed in the access point controller, or even a single access point if it included an IP packet handler.

Moreover the flow classification used in the preferred embodiment to detect an IP flow is not essential to the invention. If IP flow detection is required, then various other criteria can be used. For example, the flow classifier can be dynamically configured by changing the value of the packets/sec detection criteria parameter.

The present invention includes any novel feature or combination of features disclosed herein either explicitly or any generalisation thereof irrespective of whether or not it relates to the claimed invention or mitigates any or all of the problems addressed.

In view of the foregoing description it would be evident to a person skilled in the art that various modifications may be made within the scope of the invention.

Claims

1. A method for supporting the quality of service in packet data transmission in a radio network, whereby transmission over the air interface is in radio flows, the method comprising:

selecting a radio flow having appropriate quality of service characteristics for the packet to be transmitted over the air interface from a selection of predefined default radio flows having different quality of service characteristics.

2. A method as claimed in claim 1, wherein selecting the radio flow comprises providing the packet to be transmitted with a radio flow identifier selected from predefined default radio flow identifiers representative of different quality of service characteristics.

3. A method as claimed in claim 2, further comprising mapping the packet into the identified default radio flow for transmission over the air interface.

4. A method as claimed in any preceding claim, comprising:

detecting handover of a mobile communications device having an active connection from one radio subnetwork to another;

performing default radio flow selection for the active connection in response to handover detection.

5. A method as claimed in any preceding claim, further comprising:

monitoring packets to be transmitted over the air interface to detect IP flows;

switching a detected IP flow to a dedicated radio flow having corresponding quality of service characteristics.

6. A method as claimed in claim 5, wherein switching the detected IP flow to a dedicated radio flow comprises:

providing the packets of a detected IP flow with an identifier of the dedicated radio flow; and

mapping the packets of the detected IP flow into the identified dedicated radio flow for transmission over the air interface.

7. A radio access system for supporting the quality of service in data packet transmission over the air interface, the system comprising:

a selection of predefined default radio flows having different quality of service characteristics; and

means for selecting a radio flow having appropriate quality of service characteristics for the packet to be transmitted over the air interface from the selection.

8. A system as claimed in claim 7, wherein the radio flow selecting means comprises:

means for providing the packet to be transmitted with a radio flow identifier selected from identifiers corresponding to the predefined default radio flows.

9. A system as claimed in claim 8, further comprising means for mapping the packet into the identified default radio flow for transmission over the air interface.

10. A system as claimed in any of claims 7 to 9, further comprising means for detecting handover of a mobile communications device having an active connection from one radio subnetwork to another; and wherein the selection

means selects a default radio flow for the active connection in response to handover detection.

11. A system as claimed in any of claims 7 to 10, further comprising:

means for monitoring packets to be transmitted over the air interface to detect IP flows;

means for switching a detected IP flow to a dedicated radio flow having corresponding quality of service characteristics.

12. A system as claimed in claim 11, wherein the switching means comprises:

means for providing the packets of a detected IP flow with an identifier of the dedicated radio flow; and

means for mapping the packets of the detected IP flow into the identified dedicated radio flow for transmission over the air interface.

13. A communication device for use in a system which supports the quality of service in data packet transmission over the air interface and comprises a selection of predefined default radio flows having different quality of service characteristics, wherein the device is arranged to select a default radio flow having appropriate quality of service characteristics for the packet to be transmitted over the air interface from the selection.

14. A device as claimed in claim 13, which is a mobile communication device or a mobile router.

15. A device as claimed in claim 13 or 14, for use in a system as claimed in any of claims 7 to 12.

16. A method for supporting the quality of service in packet data transmission in a radio substantially as hereinbefore described with reference to and/or as illustrated in any one or any combination of Figures 1 to 28 of the accompanying drawings.

17. A radio access system for supporting the quality of service in data packet transmission over the air interface substantially as hereinbefore described with reference to and/or as illustrated in any one or any combination of Figures 1 to 28 of the accompanying drawings.

18. A communication device for use in a system as claimed in claim 16.

Abstract

A method and system for supporting the quality of service in wireless networks

A mechanism is provided for supporting differentiated services (quality of service) in a radio network. A radio access system is provided which supports the quality of service in data packet transmission over its air interface. The system comprises a selection of predefined default radio flows having different quality of service characteristics and means (4, 52) for selecting a radio flow having appropriate quality of service characteristics for the packet to be transmitted over the air interface from the selection.

[Fig. 14]

Method supporting the quality of service of data transmission

The present invention relates to a method as set forth in the preamble of the appended claim 1 for supporting the quality of service of data transmission in wireless communication according to the Internet protocol, a system as set forth in the preamble of the appended claim 8, and a wireless communication device as set forth in the preamble of the appended claim 14.

The International Standardisation Organisation ISO has developed an open system interconnection (OSI) model for describing the distribution of data transmission in different layers. The layers are, listed from top downwards, an application layer, a presentation layer, a session layer, a transport layer, a network layer, a data link layer, and a physical layer. In view of the present specification, the most essential layers are the physical layer, the data link layer and the application layer.

The European Telecommunication Standards Institute ETSI has defined a standard for a wireless local area network (ETS 300 652), HIPERLAN Type 1 (high performance radio local area network) to be applied *e.g.* in wireless local area networks of short distances, such as local area networks of offices. In a local area network according to this standard, several devices may be connected which communicate on the same data transmission channel using packet data transmission. The standard defines the two lowermost layers of said OSI model: the physical layer and the data link layer.

The Conference of European Post and Telephone Administrations CEPT has defined a standard TR 22-05 where the frequency range from 5.15 GHz to 5.3 GHz is reserved for data transmission according to the HIPERLAN standard. This frequency range is divided into five channels, each of which being allotted a band width of ca. 23.5 MHz. Figure 1a shows a reduced example of such a local area network according to the HIPERLAN standard. It consists of terminal nodes 101a, 101b, 101c, 101d, a switching node 102 and a gateway node 103. The terminal nodes 101a—101d may communicate directly with each other, or they may communicate via the switching node 102 if there is no di-

This Page Blank (uspto)

rect radio communication between the terminal nodes 101a—101d due to *e.g.* too long a distance or obstacles dampening radio signals. Via the switching node 102, the terminal nodes 101a—101d can also communicate with the gateway node 103 which is coupled to *e.g.* a wireless local area network 104 or the Internet network. Thus, the terminal node 101a—101d can be used as an Internet host, if necessary.

Figure 1b shows the structure of a data transfer packet according to the HIPERLAN standard. First, there is a header which is transmitted at a lower bit rate (LBR) than the other blocks and which includes the address information and the length of the packet. This is followed by a synchronisation block for synchronising the receiver to the data blocks of the packet DB(1), DB(2), ..., DB(m) containing the actual information to be transmitted. One packet may contain a maximum of 47 data blocks. Each packet can be addressed to either one receiver (unicast packet) or several receivers (multicast packet). As the third packet type the HIPERLAN standard defines an acknowledgement packet (ACK) by which the receiver of the packet informs about the successful receipt of the packet so that the sender will know if there is a need to retransmit the packet. In packets requiring data transmission in real time, it can be defined that the receipt of the packet is not acknowledged, because the information contained in the packet could be outdated if retransmitted. Packets of this kind are, for instance, packets for audio applications. On the other hand, for some real-time applications with higher quality demands, such as video applications, it is possible to define limited packet acknowledgement, whereby the acknowledgement is transmitted for several packets with one message. In packets not requiring real time, it is possible to define the acknowledgement to be sent after the receipt of each packet.

The transmission and receipt take place on the same channel without external synchronisation. The channel is listened to by the receiver of the transmitting node for a certain time, and if no communication is detected on this channel within this time, it is assumed that the channel is free and transmission is started. However, if communication is detected on this channel, the receiver is synchronised with this transmission. After the transmission, a possible acknowledgement message is waited for, and after this, an attempt for obtaining the channel can be started.

This Page Blank (uspto)

However, there may be several nodes waiting for transmission turns, whereby it may occur that several terminal devices try to transmit simultaneously. This can be solved *e.g.* so that the nodes are allotted different priorities, whereby a node with a lower priority will wait a longer time after the end of a transmission before it starts to transmit, if no communication is detected on the channel within this time.

The term "Internet" is commonly used to describe an information resource from which information can be retrieved from a data processor, such as a personal computer (PC). The data processor communicates via a modem with a telecommunication network. This information resource is distributed world-wide, comprising several storage locations which also communicate with the telecommunication network. The Internet is made operable by defining certain data communication standards and protocols, such as TCP (transfer control protocol), UPD (user datagram protocol), and IP (Internet protocol), which are used for controlling data transmission between numerous parts of the Internet. The TCP and the UDP are involved with preventing and correcting data transmission errors in the data transmitted in the Internet; the IP is involved with data structure and routing. The currently used versions of the Internet protocol are IPv4 and IPv6.

Thanks to the growing popularity of open data systems, the Transmission Control Protocol/Internet Protocol (TCP/IP) communication protocol has become a generally used protocol whereby computers of different sizes and brands can communicate with each other. TCP/IP support is currently available for almost all operating systems. The network layer protocol of TCP/IP, the Internet Protocol IP, is intended to be routed by gateways, *i.e.* routers. The routing is conducted by means of IP addresses of four bytes and routing tables. Thanks to the Internet protocol, computers using the TCP/IP can transfer messages in the routing network even to the other side of the world.

The Internet, which covers well particularly the industrialised countries, is a huge network of routers using the TCP/IP communication protocol. The largest group of users of the Internet, which was originally in scientific use only, is now firms which buy their services from commercial connection providers. In the Internet, each device has its own individual

This Page Blank (uspto)

IP address. In the Internet protocol version IPv4, the IP address consists of 32 bits, *i.e.* it is a digit of four bytes which is divided in two parts: an organisation-specific network address and a network-specific device address. For facilitating the processing of addresses, a decimal dot notation system has been introduced, in which the addresses are indicated by digits of 8 bits separated by dots (an octet). One octet is a number from 0 to 255. This address mechanism is further divided into three different classes (ABC) which make network and device addresses of different lengths possible.

Further, with the growing popularity of the Internet, the length of the address blocks in the data packets of Internet messages is no longer sufficient in all situations for indicating all the addresses in use. This is one reason for developing the Internet protocol version IPv6. In this protocol version, the length of the address blocks is increased to 128 bits, which means in practice that an individual address can be reserved for all devices that are connected with the Internet network. Figure 2 shows the blocks of the data packet in Internet messages.

The header block consists of the following elements:

Version	IP version of 4 bits (=6)
Prio.	4 bit priority,
Flow label	24 bit label for identifying the connection in the application layer,
Payload length	16 bit integer indicating the length of the payload, *i.e.* the length of the packet after the header in bytes,
Next header	data of 8 bits determining the header immediately following the IPv6 header,
Top limit	integer counter of 8 bits which is reduced by one at the each device (node) which transmits the packet further; the packet is rejected if the value is reduced to zero,
Source address	the 128 bit address of the sender of the original packet,
Destination address	the 128 bit address of the intended recipient.

This Page Blank (uspto)

The header is followed by the payload block, *i.e.* the actual information to be transmitted.

Physically, the Internet consists of communication network arranged in a hierarchy, for example local area networks (LAN), regional telecommunication networks, and international telecommunication networks. These communication networks are coupled internally and externally with routers which transmit information from the transmitting terminal equipment or from the preceding router in the chain of data transmission, and route the information to the receiving terminal equipment or to the next router in the chain of data transmission.

Figure 3 shows the coupling of the transmitting terminal equipment (source host, SH) and the receiving terminal equipment (destination host, DH) to the Internet via corresponding local area networks LAN and routers R.

Below in this specification, the transmitting terminal equipment and receiving terminal equipment will also be called by the common term Internet host. The Internet host can be typically used either as the source host SH and the destination host DH.

An Internet host, coupled to the Internet network via a local area network LAN, is either provided with a permanently defined Internet address or the address is a dynamic address generated by the server of the local area network (for example by using a dynamic host configuration protocol DHCP). In case the Internet host is coupled by a modem to a telecommunication network, the telecommunication terminal must ask for an Internet address from an Internet service provider to which the Internet host is registered. This is conducted *e.g.* according to a point-to-point protocol (PPP) formed above the Internet protocol layer. In both cases, the information to be transmitted in the Internet is routed to the Internet host possibly via several communication networks and routers from a remote host by using a determined Internet address.

The IP defines the transmission of the communication in packets (datagrams). The packet data transmission is one reason for the popularity of the Internet, because it allows transmission in bursts which does not

This Page Blank (uspto)

require constant on-line connection and makes it possible that several Internet hosts are coupled in the same telephone connection. When a router receives a packet containing a destination address, the router routes the packet forward, if there is free capacity in the buffer memory of the router and at least one open telephone line. If there is not sufficient memory space or no open telephone line available at the moment, the packet is rejected and the source host or the preceding router must try retransmission later. In general, the Internet does not support time-critical data transmission, and the method of best effort offered by the Internet protocol is sufficient.

In the transmission of packets according to the Internet protocol, the packets can be transmitted directly to the receiver only when the network elements of the addresses of both the host and the destination are the same. In other cases, the packets are transmitted to a router which takes care of transmitting the packets further, either to the next router or to the destination, if the recipient is in the network of the router. In each router, each packet entering the router is transferred from the communication layer according to the OSI model to the network layer, where the header of the packets is examined, and on the basis of the address data therein, a decision is made where the packet is to be transmitted. For transmission, the packets are transferred back to packets of the communication layer. Because the Internet protocol has the character of a connectionless protocol, the above-mentioned operations must be taken for each packet entering the router. If the communication layer is fast, for example in accordance with the asynchronous transfer mode ATM, the processing of the packets takes a significant part of the time used for transmission. Thus, the whole transmission capacity of the transfer line cannot be utilised effectively. For correcting this situation, *e.g.* Ipsilon Networks has developed a coupling solution. In this solution, an attempt is made to detect time-consuming data transmission flows and to couple them directly with a communication layer.

The coupling solution by Ipsilon Networks consists of switches and controllers for controlling their operation. When a continuous communication flow is detected by the controller in any protocol communication in the Internet, the controller requests the transmitter to label the pack-

This Page Blank (uspto)

ets of said communication flow with a flow label, *i.e.* to open a so-called virtual channel for this communication flow. If the same finding is made by the receiver, also it requests for separation of the communication flow onto a separate virtual channel. Subsequently, this controller between the transmitter and the receiver may locally control their own switch to turn on direct communication between these two virtual channels. Because the presented coupling solution is based on labelling communication flows, it contains for each label a defined time limit after which the label is rejected, if there is no longer communication on the channel labelled by it. This reduces the number of different labels required simultaneously. In this solution, the coupling is made on the basis of communication between three nodes, and the switching request is made by the sender and/or the receiver. The coupling reduces primarily the delay of data transmission in comparison with routing.

This coupling solution is only intended for accelerating routing of packets according to the Internet protocol, and this coupling solution requires that three nodes are involved. This solution does not consider the quality of service as such.

Data transmission in packet form improves the degree of capacity utilisation of the communication channel in general, not only for retrieving information from the Internet. For example, packet data transmission can be used in applications, such as voice calls, video negotiations and other communications according to different standards. However, some of these applications are time-critical. For example in a real-time voice call, the service of best effort offered by the Internet protocol may cause significant delays in the transmission and transfer of the audio signal, which affects the understanding of the received audio signal so that *e.g.* speech is almost or totally intelligible. Moreover, the delay (the time consumed from the transmission to the receipt of the packet) may vary during the transmission of the audio signal, depending on *e.g.* the load of the communication network and variations in transmission errors. The same applies also to the transmission of a video signal in real time. There may also be situations where the users of Internet do not want as long delays as occur in many cases for obtaining information from the Internet.

This Page Blank (uspto)

The Internet Engineering Task Force (IETF) is an organisation involved with the development of Internet architecture and operation in the Internet. The IETF is currently developing a new protocol which provides an Internet host the possibility to request a desired quality of service from available defined qualities of service (QoS). This protocol is known as the resource reservation protocol (RSVP), and it is presented in the standard proposition "Resource ReSerVation Protocol (RSVP) - Version 1 Functional Specification" by Braden, R.; Zhang, L.; Berson, S.; Herzog, S.; Jamin, S.; RFC 2205, September 1997 (available at http://www.isi.edu/div7/rsvp/pub.html). The Internet host uses the RSVP protocol when requesting a certain quality of service QoS from the Internet network on the basis of the communication flow of the application which the Internet host wishes to receive from a remote host. The RSVP protocol transmits the request through the network by using each router user by the network for transmitting the communication flow to the receiving Internet host. In each router, the RSVP protocol tries to make a resource reservation for said communication flow. Also, the RSVP protocol tries to make a resource reservation for the communication flow in the destination and source Internet host.

For making a resource reservation in any node, which may be either a router or an Internet host, the RSVP protocol communicates with two local terminal modules: access confirmation module and protocol module. The access confirmation module deduces whether the node has sufficient resources for providing the requested quality of service. The protocol module deduces if the user has access to make a reservation. If either checking fails, the RSVP protocol returns an error message to the application that formed the request. If both tests are successful, the RSVP protocol sets the parameters for classifying the packet and for scheduling the packet in the Internet source host for achieving the desired quality of service. The classification of the packet deduces for all packets a service quality class, and the scheduling controls the transmission of the packets for achieving the promised quality of service in all communication flows.

The RSVP protocol operates on top of the Internet protocol both in IPv4 and IPv6. In particular, the RSVP protocol is designed to utilise the strong points of the routing algorithms of the present Internet. The

This Page Blank (uspto)

RSVP itself does not conduct routing but it uses routing protocols of lower levels to deduce where reservation requests should be transferred. Because the routing changes the routes for complying with changes in the topology of the Internet network, the RSVP protocol places its reservations for resources in new routes, if necessary.

Telecommunication networks and the Internet are two significant worldwide communication networks, whereby wireless telecommunication terminals are developed for coupling therewith and for their use. For example, cellular networks make it possible to couple a wireless telecommunication terminal to a telecommunication network and offer a high quality of service with circuit-switched technology. These cellular networks and other mobile communication networks can be utilised also for coupling to the Internet network and for utilising multimedia services. However, the circuit-switched system has the disadvantages that, the connection from a wireless telecommunication terminal to a wireless communication network is turned on during the whole connection, which takes up the capacity of the wireless communication network and limits the number of simultaneous connections.

In solutions of prior art for wireless packet communications, obtaining a quality of service is not supported. Because of this, a development in the Internet community has been started for solutions supporting the mobility of Internet host and obtaining quality of service in the Internet protocol version IPv6.

In radio links, data is typically transmitted in a channel which is a certain frequency range. In one system, several channels can be available simultaneously. Further, in full duplex data transmission there are separate transmitting and receiving channels, whereby for example a base station transmits on the transmitting channel to the terminal device and the terminal device transmits on the receiving channel to the base station. A problem with radio links is that the radio channel is a limited resource which limits *e.g.* the band width and/or number of channels that can be reserved as well as the data transmission rate available for the radio link. The radio channel is liable to disturbances, such as distortion of the received signal caused by multi-channel propagation which is due to the fact that the same signal is received at

This Page Blank (uspto)

the destination through different routes at different times. To reduce the effect of disturbances, part of the data transmission capacity must be used for transmitting error correction data with the packets, and achieving a desired error probability rate may require several packet retransmissions, which reduces the capacity of the radio link.

In radio links where several data transmission flows are transmitted on one channel, packets of these different data transmission flows are multiplexed. The transmission order can be affected by arranging packets of different data transmission flows in an order of priority, whereby packets of a flow with higher priority are transmitted more often than packets of a flow with lower priority. These include packets of a real-time application which are preferably made as short as possible. On the other hand, packets of applications with lower priority are often considerably longer than packets with higher priority. In systems of prior art, such a long packet prevents the transmission of other packets as long as the transmission of the packet takes. This may cause considerable delays also in the transmission of packets with higher priority, and reduce the quality of service.

It is an aim of the present invention to provide a method for flexible determination of the quality of service in wireless communication in the Internet. The method of the invention is primarily characterised in what will be presented in the characterising part of the appended claim 1. The system of the invention is primarily characterised in what will be presented in the characterising part of the appended claim 8. Further, the wireless communication device of the present invention is primarily characterised in what will be presented in the characterising part of the appended claim 14. The invention is based on the idea that for setting up an Internet connection, the required quality of service is determined for the connection, on the basis of which the connection is attempted to make in a wireless communication network with parameters complying with the set quality of service.

The present invention gives significant advantages to the solutions of prior art. In a wireless connection set up by the method of the invention, the quality of service is obtained in a more reliable way, and moreover, the whole capacity of the wireless communication network can be util-

This Page Blank (uspto)

ised more efficiently, because for some connections it will suffice to have a quality of service which takes up less of the capacity of the communication network. On the other hand, fewer retransmissions will be required in connections where no high demands are set for the correctness of the data transmission, *e.g.* for the transmission of speech or video between the Internet network and a wireless telecommunication terminal. Thus, more capacity will be left for applications where *e.g.* the correctness of data transmission is important, such as in the transmission of data files. Data transmission flows are examined in a centralised manner on the Internet protocol level, and a detected data transmission flow is supplemented with the parameters of the quality of service of the radio interface. These parameters are obtained advantageously from a predetermined data file depending on the radio interface. In this centralised definition of quality of service, only two parties are required, and it is possible to better examine active data transmission flows and quality of service levels defined for them, before setting the quality of service for a new data transmission flow. Thus, new connections will not reduce the quality of service of existing connections.

As another advantage, it can be also mentioned that the packet of one data transmission flow does not need to be transmitted as one packet but it can be divided into smaller parts which are, according to the invention, equipped with a label of the radio flow, on the basis of which the receiver can distinguish between packets of different flows and their parts. Thus, between parts of one long packet, it is possible to transmit a packet of a flow requiring higher quality of service. Further, the number of retransmissions can be reduced, because errors occur typically in bursts, whereby not all parts of a long packet are not necessarily erroneous and these do not need to be retransmitted.

In the following, the invention will be described in more detail with reference to the appended drawings, in which

Fig. 1a shows an example of a local area network complying with the HIPERLAN standard in a reduced manner,

Fig. 1b illustrates the structure of a data transmission packet complying with the HIPERLAN standard,

This Page Blank (uspto)

Fig. 2 illustrates the structure of a packet of the Internet protocol version IPv6,

Fig. 3 shows the coupling of a transmitting and receiving host via the Internet network in a reduced chart,

Fig. 4 shows the coupling of a wireless Internet host to the Internet network in a reduced chart,

Fig. 5 shows packet data transmission between a wireless communication device and the Internet network via the GSM cellular network,

Fig. 6 shows an example of generating a radio flow label upon detecting a data transmission flow in a wireless communication network, and

Figs. 7a and 7b show examples of packet transmission sequences according to prior art and upon transmission with a radio flow label according to the invention.

In the following, the invention will be described by using the GSM cellular network as an example of a wireless communication network and a wireless communication device of the GSM system as the wireless Internet host, but the invention can also be applied in other wireless communication networks and wireless telecommunication terminals with the option for data transmission in packets. This wireless communication device 1 can also consist of a computer, such as a portable computer, coupled with a wireless data transmission device, such as a radio modem.

In this specification, data flow refers to the transmission of data packets belonging to the same communication/application. Respectively, wireless data flow refers to the transmission of data packets belonging to the same communication/application, advantageously via the radio channel, whereby also the term radio flow is used. The packets may be *e.g.* packets complying with the Internet protocol or GPRS packets of

This Page Blank (uspto)

the GSM cellular network. The GPRS packet transmission system provides the possibility of 14 simultaneous connections in one terminal (wireless communication device) at the data transmission level. At present, the GPRS packet transmission offers the possibility of arranging the packets in four different levels of priority. The block reserved for this priority information in the packet can be modified into a block reserved for the radio flow label according to this invention. In case there is a need to form at least as many radio flows as the number of simultaneous connections, the corresponding number of bits are reserved for the radio flow label. Thus, two additional bits will be needed in addition to the priority block.

Each connection may be connected with one application, but the same application may involve also more than one connection. The data transmission flows of these different connections belonging to the same application can be identified by the address and port data of the sender and the receiver in the header of the packets.

Figure 4 is a reduced chart showing the coupling of a wireless Internet host in the Internet network. The system consists of a wireless communication device 1, a radio access network 2 and a core network 3. The radio access network comprises the operations for accomplishing data transmission between the wireless communication device 1 and the core network 3 as well as for controlling wireless resources, for setting up and down wireless data flows or radio flows, for moving the connection from one control station to another (handover), and possibly also for compressing packets *e.g.* according to the IPv6 standard. In this example, the functional elements of the radio access network include an access point 4, 4' (AP) and an access point controller 5 (APC). A radio communication is set up between the access point 4 and the wireless communication device 1, for transmitting *e.g.* signals required for setting up the connection and information during the connection, such as data packets of an Internet application. The access point controller 5 controls over one or several access points 4, 4' and connections set up through them to wireless communication devices 1. The radio access network 2 may comprise several access point controllers 5, 5', 5". In the GSM cellular network, the access point 4, 4' is a

This Page Blank (uspto)

base station and the access point controller 5, 5', 5" a base station controller.

The core network consists of nodes connected by wires in the Internet, such as routers and wired Internet hosts.

The core network can be divided into so-called domains. These domains have a server computer or a corresponding router, by means of which the domain can communicate with other domains in the Internet. The Internet hosts in the domain, in turn, are coupled with the router of the domain. Figure 4 shows a core network with two such domains 6, 6' which are intended for serving wireless communication devices 1. These domains 6, 6' include mobile domain (MD) routers 7, 7' which control the access point controllers 5, 5', 5" coupled with the domain 6, 6'. Mobility is achieved in Internet protocol version 6 by supplementing the protocol with a data transmission method whereby the domains can transmit information from a wireless Internet host that has changed its domain. This data transmission method is called in this specification a home agent. In this context, reference is made to the Internet protocol standard version 6 IPv6, where operation of this home agent is described in more detail. The mobile domain router 7, 7' contains the functional properties of the dynamic host configuration protocol version 6 DHCPv6 and the monitoring of the mobility of the wireless communication device 1 between the access point controllers 5, 5', 5" coupled within the mobile domain 6, 6'. It should be mentioned that in some domains, there may be one or several conventional routers between the mobile domain router 7, 7' and the access point controller 5, 5', 5", even though these possible routers are not shown in the appended Fig. 4. In the GSM cellular network, where the general packet radio service GPRS is used, the element corresponding to the mobile domain router 7, 7' is the serving GPRS support node SGSN. The element corresponding to the home agent in said GSM cellular network is the gateway GPRS support node GGSN.

The network architecture used as an example in this specification gives an outline on how the quality of service can be defined in band-limited radio access networks when coupled with the Internet network. This architecture involves two connection interfaces: the radio interface and

This Page Blank (uspto)

the radio access network / core network interface. Thus, the radio interface is generated for communication between the wireless communication device 1 and the access point 4, 4'. In a corresponding manner, the radio access network / core network interface consists of the connection between access point controllers 5, 5', 5" and mobile domain routers 7, 7'.

The user of the wireless communication device 1 can use the Internet network *e.g.* in a way that an application program, such as a browser, designed for this purpose is turned on in the wireless communication device 1. In the application program, the user of the wireless communication device sets as the destination address the address of a desired Internet server or Internet host, for example the address of the Internet server of the provider of the service with which the user of the wireless communication device has made a subscription to using Internet services. As already presented above in this specification, this Internet address can be given as a four-part octet number string or addresses in text form can be used, whereby a domain name server converts the address from text form into a numerical string according to the Internet protocol.

Figure 5 is a chart showing a situation where the wireless communication device 1 is coupled to the Internet network via a digital cellular network by using the general packet radio service GPRS. The wireless communication device 1 communicates with an access point 4 on any channel of the frequency range reserved for the system. In the GSM cellular network, this access point 4 is a base transceiver station (BTS) of the base station subsystem (BSS). One access point 4 forms the radio interface of one cell in the cellular network. The access point 4 operates as a transmitter of information to be transmitted between the wireless communication device 1 and the access point controller 5. It is a central function of the access point controller to control the channels in the interface and to transmit the connection from one access point 4 to another access point 4' in a situation when the wireless communication device 1 moves from one cell to another.

Next, data transmission from another Internet host to the wireless communication device 1 will be described. The Internet application of

This Page Blank (uspto)

the wireless communication device 1, to which the information is finally transferred, transmits the above-mentioned address to define the source Internet host. The data transmission is conducted according to the GPRS standard from the mobile station 1 to the GSM cellular network. The GSM cellular network converts the packet message to a message complying with the Internet protocol and transmits it to the Internet network. The information formed in the application is transmitted to the wireless communication device 1 according to the Internet protocol via the Internet network in a manner known as such by routing to the GSM cellular network, where the information is converted to comply with the packet transmission mechanisms of the cellular network, in this case into packets of the GPRS network. The information is transmitted further via the access point controller 5 to the access point 4 and further to the wireless communication device 1 where the received message is transferred to the application layer to be processed by the application.

The following is a description on the method according to an advantageous embodiment of the invention for generating a radio flow label in communication between the wireless communication device 1 and the access point 4, 4'. The application is an Internet application of the wireless communication device 1, from which information complying with the Internet protocol is transmitted to the Internet network. This specification does not contain a more detail description on the formation of packets between the wireless communication device 1 and the mobile communication network, which may vary in different mobile communication networks and is prior art known as such by an expert in the field. Figure 6 is a schematic diagram of this formation of the radio flow label for data transmission between the wireless communication device 1 and the access point controller 5. All data transmission is based on packets and is routed according to the Internet protocol. The mobile terminal radio flow agent (MRFA), which is implemented advantageously in the application software of the wireless communication device 1, starts to transmit radio flow information packets using a default flow ID. At the access point 4, an access point radio flow agent (ARFA) transmits the flow further to the access point controller 5. At the access point controller 5, a router matrix (RM, not shown) transmits the flow to a radio flow manager block (RFM). The access point controller 5

This Page Blank (uspto)

detects that this flow is of the kind for which a radio flow label should be formed for achieving a certain quality of service (block 601 in Fig. 6). The access point controller 5 finds out if there are sufficient resources available at the moment to be used for data transmission between the wireless communication device 1 and the access point 4 in order to achieve the desired quality of service for said flow FID (block 602). If sufficient resources are available, the radio flow manager RFM selects a new flow label for the flow to be transmitted via the access point 4 to the mobile terminal radio flow agent MRFA of the wireless communication device 1. In the selection of the flow label, TCP/IP ports and/or addresses of the source host and the destination host are used. This flow label is for example data of 20 bits transmitted to the wireless communication device 1 via the access point 4. In Fig. 6, this step is indicated by arrow 603, and although it is connected directly from the access point controller 5 to the wireless communication device 1, in practical applications it is transmitted physically via the access point 4. In the wireless communication device 1, this received flow label is processed, and on the basis of this, the wireless communication device 1 generates a shorter flow label, in this application example a flow label of 8 bits, wherein a total of 256 different flow labels can be used simultaneously for different Internet applications in one wireless communication device 1.

The access point controller 5 transmits the same flow label also to the access point 4 (arrow 604); in addition, information can be transmitted here on what kind of a quality of service is desired for this flow.

The shorter flow label generated in the wireless communication device 1, which in this specification will be called the radio flow identification (RFID), is transmitted from the wireless communication device 1 via the radio interface MT/RP of the wireless communication device to the access point 4. As known, each wireless communication device of the cellular network is equipped with a device identification or a corresponding separate identification whereby wireless communication devices of the cellular system can be separated from each other. The radio interface MT/RP of the wireless communication device includes, in a manner known as such, a radio transceiver (not shown) as well as coding/decoding means (not shown), but it will not be necessary to de-

This Page Blank (uspto)

scribe this radio interface in more detail in this context. This mobile station identification MSID, which in the GSM system is advantageously the international mobile equipment identity IMEI, is transmitted from the wireless communication device 1 to the access point 4 in connection with the transmission of messages (arrows 605 and 606). Now, the access point 4 has the flow identification FID, the radio flow identification RFID as well as the mobile station identification MSID. After this, on the basis of radio flow identifications RFID coming from the wireless communication device and the mobile station identification MSID, the access point 4 can couple the flow with the original wider flow identification FID. The access point 4 transmits an acknowledgement message to the wireless communication device 1 (arrows 607 and 608) and to the access point controller 5 (arrow 609). After this, also the wireless communication device sends an acknowledgement to the access point controller 5 (arrow 610). Now, there is a connection corresponding with the desired quality of service between the wireless communication device 1 and the access point controller 5 (this is shown by block 611).

Also, the access point controller 5 may receive from the Internet network a data flow addressed to the Internet application of the wireless communication device 1. Thus, the access point controller 5 finds that a flow label can be defined for this flow, whereby the access point controller 5 examines the quality of service desired for the flow and finds out if there are sufficient resources available for achieving and maintaining the desired quality of service. At this point, the access point controller 5 considers also the other radio flows active at the moment and finds out if the desired quality of service can be provided for this flow without risking the quality of service of the active flows. If the quality of service can be achieved, the above-mentioned signalling is conducted, whereby *e.g.* a flow ID is defined for the radio flow.

In case there are no sufficient resources available on the radio channel for achieving the desired quality of service, it is possible *e.g.* to continue the radio flow at a level with a poorer quality of service, for example with a transmission of best effort, whereby the source host of the flow is informed of this procedure. If necessary, the user can be inquired if the data is to be transmitted in spite of the lower quality of service or if the data transmission is to be interrupted.

This Page Blank (uspto)

The information transmitted from the second host according to the Internet protocol is transmitted via normal mechanisms of the Internet protocol to the cellular network. In the cellular network, the message is converted to a message corresponding with the packet transmission mechanisms of the cellular network and transmitted to the access point controller 5. The access point controller 5 provides the message with a flow identification FID and transmits the message further to the access point 4. At the access point 4, it is examined on the basis of this flow identification FID what are the corresponding radio flow identification RFID and mobile station identification MSID. Next, the flow identification FID is removed at the access point 4 and replaced by the radio flow identification RFID. This way it is possible to reduce the information to be transmitted along with the packets (in this example 20 — 8 = 12 bits), which reduces the load of the radio network and makes it possible to utilise the radio network more efficiently. This is also illustrated in the appended Fig. 7a showing four transmission strings 701, 702, 703, 704 containing packets of radio flows. As examples, the packets of each string are indicated by the number of the connection (1 to 7) to which the packet belongs. Of these strings, the access point controller 5, 5', 5" selects the packet to be transmitted at each time on the basis of predetermined criteria. Prior art is shown by the first transmission sequence 705 where the order of transmission is determined primarily on the basis of priority set for the string. In this example, the order of priority is the following: the highest priority belongs to the string 701, next to the second string 702, third to the string 703, and the lowest priority to the string 704. Header blocks are indicated by letters H in each packet.

Data transmission according to an advantageous embodiment of the invention is illustrated by the second transmission sequence 706. In this situation, the transmission order of the strings 701 to 704 is determined according to the quality of service set for the radio flow corresponding to the string in a way that the higher quality of service is set for the first string 701, the next highest to the second string 702, next to the third string 703, and the lowest quality of service is set to the fourth string 704. The radio flow identifications are indicated in this second transmission sequence 706 with the reference numeral 707.

This Page Blank (uspto)

The wireless communication device 1 receives a packet message according to this transmission sequence and transmits the information contained in it to the corresponding application. The wireless communication device 1 contains also a switching table or the like containing information on the application to which a certain radio flow identification RFID corresponds. Also transmission from the wireless communication device 1 to the Internet network is conducted in a reverse order, applying the same principle.

In the formation of the packet transmission sequence, it is possible to consider *e.g.* the number of strings 701 to 704, retransmission needs caused by errors, statistical multiplexing for packets of fixed size, an attempt to reduce the average delay, and utilisation of the channel as efficiently as possible.

For defining the quality of service QoS, it is possible to utilise information in the header of the application received in the Internet message. At the present, a standard is under development on how these qualities of service could be presented and what they could be. In any case, a message according to the Internet protocol contains, in the header, information about the type of the application, which can be *e.g.* an audio application, a video application, a data application, or a combination of these. These applications of different types have different requirements. For example, the real-time processing of audio and video applications usually requires that the packets must be transmitted to the destination within a certain response time or otherwise the packets must be rejected. However, in data transmission, for example in the transmission of program files, it is the correctness, not real-time processing, of data transmission that is important. In presently known methods and cellular networks, it is defined at the design stage, what is the error probability of data transmission, on the basis of which it is possible to select error correction algorithms and to set *e.g.* a maximum number of retransmissions. All packet information is transmitted according to the same criteria. If any packet is transmitted incorrectly, it is retransmitted. These retransmissions are conducted either as long as the packet is received correctly or, if a response time is defined for the packet, the packet is rejected if it cannot be received within the pre-

This Page Blank (uspto)

scribed time or the maximum number of retransmissions is exceeded. Since in audio and video applications even a partly incorrectly received information would be sufficient, this retransmission constitutes an unnecessary load on the radio network. On the other hand, the additional load reduces the radio resources available for other applications and thus interferes also with the quality of service obtained by other applications. For detecting and correcting errors, several methods have been developed which are prior art to an expert in the field, wherein it is rendered unnecessary to discuss them in more detail in this context. It should be further mentioned that increasing error detection and error correction capacity by error detection and correction algorithms will increase the need of data transmission. These conflicting demands set a limit to the fact how efficient an algorithm is selected, to prevent an unnecessary delay in the data transmission.

When using a method of the invention, it is possible to define different qualities of service with different demands. For example, a poorer error probability demand can be defined for audio and video packets than for data packets. On the other hand, due to the real time demand, a higher priority can be determined for audio and video packets than for data packets. Thus, data packets are transmitted at a slower rate, if the radio network is loaded. Further criteria describing the quality of service may include response time, within which the packet must be received or else it is rejected. By combining these different criteria, several different qualities of service are obtained, and also other criteria than those mentioned above can be used in defining the quality of service.

These qualities of service and the corresponding bits of the header to be examined are *e.g.* listed in a table by the access point controller 5, whereby by examining these header bits, the access point controller 5 retrieves the corresponding quality of service from the table. For these qualities of service, information is stored in the access point controller 5 on the special demands of each quality of service, including the above-mentioned error probability, priority and response time.

These definitions for the quality of service are transmitted from the access point controller 5 to the access point 4 which, on the basis thereof, conducts the definition of the transmission order of the packets to be

This Page Blank (uspto)

transmitted. There may be several Internet applications to be transmitted by one access point 4 simultaneously. For these different applications, a string is preferably formed for each, where packets are transferred for transmission. From these packets in different strings, the access point 4 selects the packet to be transmitted at the time.

According to the invention, it is possible to use the radio flow label to improve the efficiency of the system also in a way that the transmission of long packets can be divided into parts so that, if necessary, one or several packets of a higher quality of service are transmitted between the parts. Such a part can be *e.g.* in a time-division radio link one time period. In systems of prior art, the whole packet must be transmitted in subsequent time periods, because the receiver cannot otherwise identify the flow to which the packet part belongs. In the system of the invention, the packet parts can be identified on the basis of the radio flow identification. This situation is illustrated in the appended Fig. 7b showing four strings. Each string contains one or more packets to be transmitted. The transmission of prior art is illustrated in the first transmission sequence 705, and the transmission of packets equipped with a radio flow identification according to the invention is illustrated in the second transmission sequence 706. Thus, retransmission of so many time periods will not be needed, because, instead of retransmitting the whole packet, only the incorrectly received part or parts of the packet are retransmitted.

Determination of the quality of service according to the invention can be used also in other packet data transmission protocols and information networks. Also, in addition to the routings known from Internet networks, the invention can be applied in coupling solutions developed for Internet networks where the router is used for examining the route between data flows and conducting the coupling in the hardware layer.

The method described above as the method supporting the quality of service is applicable also together with the Internet resource reservation protocol RSVP. Thus, in the access point controller 5, 5', 5" which monitors data transmission flows, it is possible to consider also the data contained in the data transmission flow about the quality of service presented by the host. The radio flow manager block RFM formed in the

This Page Blank (uspto)

access point controller 5, 5', 5" stores the parameters of the quality of service requested by the host and finds out whether the requested quality of service is available. If the requested quality of service is available, it is possible to set the parameters corresponding to the desired quality of service for the data flow in question.

The invention is not limited solely to the embodiments presented above, but it can be modified within the scope of the appended claims.

This Page Blank (uspto)

Claims:

1. A method for supporting the quality of service (QoS) in packet data transmission between a wireless communication device (1) communicating with a radio network, and an information network (LN), where data transmission between the wireless communication device (1) and the radio network (2) is controlled with at least one access point controller (5, 5', 5"), and in which method information is transmitted between the wireless communication device (1) and the access point controller (5, 5', 5") in radio flows, **characterised** in that in the method, at least one radio flow is provided with a defined radio flow identification (RFID) and a quality of service (QoS).

2. The method according to claim 1, **characterised** in that the quality of service (QoS) is determined in a centralised manner, preferably by the access point controller (5, 5', 5").

3. The method according to claim 1 or 2, **characterised** in that for determining the quality of service (QoS), the content of the packets, preferably the content of the header (H) of the packets, is used.

4. The method according to claim 3, **characterised** in that the data transmission is divided at least into a network layer and a physical layer, wherein in the method, the data transmission is conducted in packets of the network layer, which are converted into packets of the physical layer to be transmitted in a radio flow, and that the quality of service (QoS) is determined on the basis of the contents of the packets of the network layer.

5. The method according to any of the claims 1 to 4, **characterised** in that the packets of the radio flow are formed from packets complying with the Internet protocol.

6. The method according to any of the claims 1 to 5, **characterised** in that the packets of the radio flow are transmitted in the radio network (2) as GPRS packets.

This Page Blank (uspto)

7. The method according to any of the claims 1 to 6, **characterised** in that the method comprises the steps of:

— transmitting several different radio flows in packet data transmission between the wireless communication device (1) and the radio network (2), and

— transmitting a packet of a second radio flow between packets of a first radio flow.

8. A system for supporting the quality of service (QoS) in packet data transmission in a radio network (2), the system comprising:

— at least one wireless communication device (1) communicating with the radio network (2),

— means (7, 103, GGSN) for transmitting information between the radio network (2) and the information network (3, LN),

— means (5, 5', 5") for controlling data transmission between the wireless communication device (1) and the radio network (2), and

— means (4, 4', 102) for transmitting information between the wireless communication device (1) and the access point controller (5, 5', 5") in radio flows,

characterised in that the system comprises further:

— means (5, 5', 5", 103, RFM) for determining a radio flow identification (RFID) for at least one radio flow, and

— means (5, 5', 5") for determining the quality of service (QoS) for the radio flow.

9. The system according to claim 8, **characterised** in that it comprises means (5, 5', 5") for determining the quality of service (QoS) in a centralised manner.

10. The system according to claim 8 or 9, **characterised** in that it comprises means (RFM) for determining the quality of service (QoS) on the basis of the contents of the packets, preferably the contents in the header (H) of the packets.

11. The system according to claim 8, 9 or 10, **characterised** in that it comprises means (7, 103) for generating packets of a radio flow from packets complying with the Internet protocol.

This Page Blank (uspto)

12. The system according to any of the claims 8 to 11, **characterised** in that it comprises means (GGSN, SGSN) for conducting data transmission in the radio network (2) in GPRS packets.

13. The system according to any of the claims 8 to 12, **characterised** in that it comprises:

— means for transmitting at least a first and a second radio flow in packet data transmission between the wireless communication device (1) and the radio network (2), and

— means (5, 5', 5") for transmitting a packet of the second radio flow between packets of the first radio flow.

14. A wireless communication device (1) equipped with means for transmitting information into a radio network (2), comprising:

— means (7, 103, GGSN) for transmitting information between a radio network (2) and an information network (3, LN),

— means (5, 5', 5") for controlling data transmission between the wireless communication device (1) and the radio network (2), and

— means (4, 4', 102) for transmitting information between the wireless communication device (1) and the access point controller (5, 5', 5") in radio flows,

characterised in that the wireless communication device (1) comprises further:

— means (MRFA) for generating a radio flow identification (RFID) for at least one radio flow, and

— means (MRFA) for connecting said radio flow identification (RFID) into packets of said radio flow transmitted from the wireless communication device (1).

This Page Blank (uspto)

Abstract

The invention relates to a method for supporting the quality of service (QoS) in packet data transmission between a wireless communication device (1) communicating with a radio network, and an information network (LN), where data transmission between the wireless communication device (1) and the radio network (2) is controlled with at least one access point controller (5, 5', 5"). Further, in the method, information is transmitted between the wireless communication device (1) and the access point controller (5, 5', 5") in radio flows. In the method, at least one radio flow is provided with a defined radio flow identification (RFID) and a quality of service (QoS).

Fig. 1a

This Page Blank (uspto)

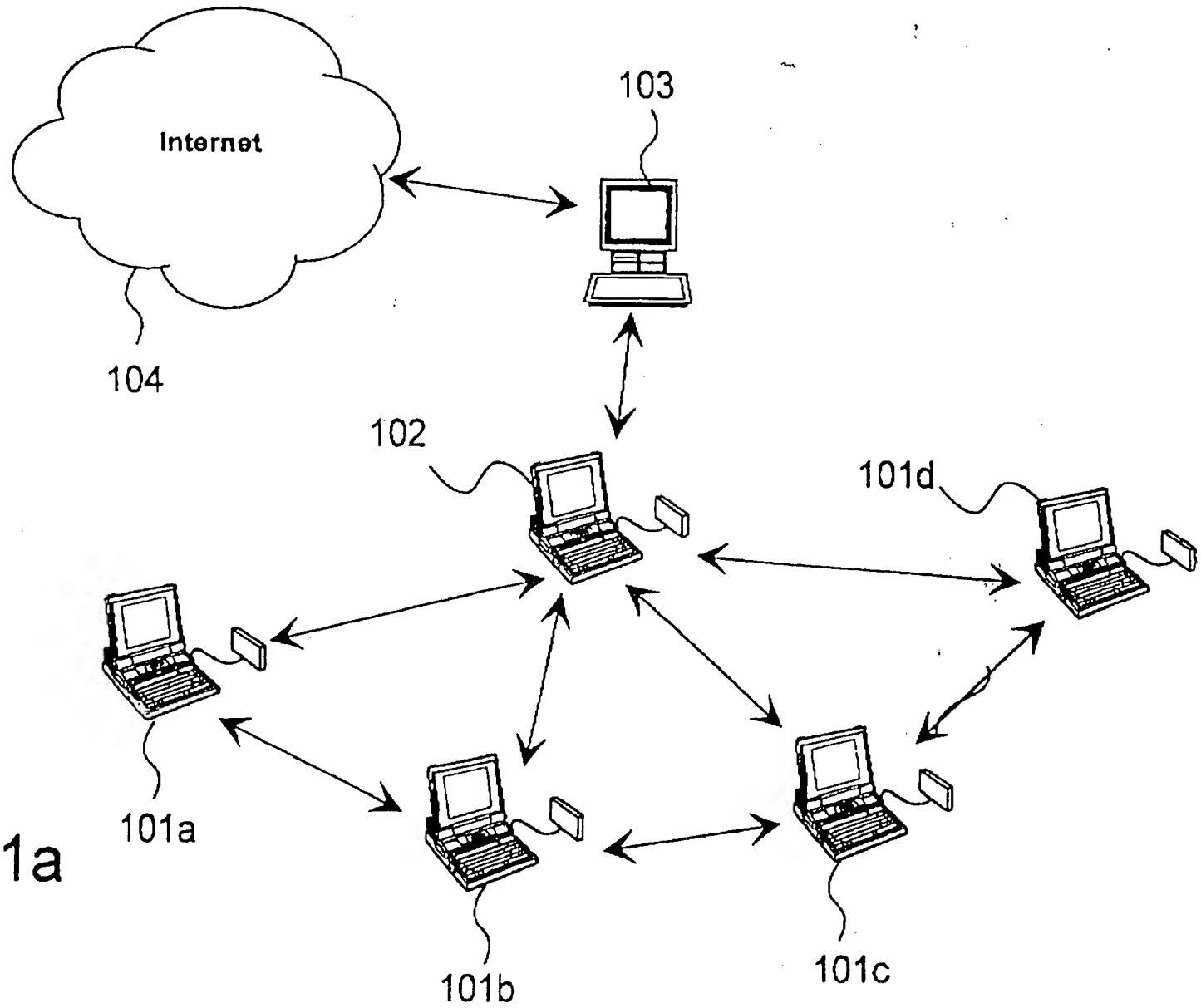


Fig. 1a

This Page Blank (uspto)

LBR Header + physical data layer data packet

8 bit Clock and Sync.	8 bit Address	2 bit Protection	6 bit Packet length	2 bit Protection	496 bit Data block 0..46

LBR Acknowledgment

8 bit Clock and Sync.	8 bit Return Address	2 bit Protection	6 bit Acknowledgment	2 bit Protection

Fig. 1b

This Page Blank (uspto)

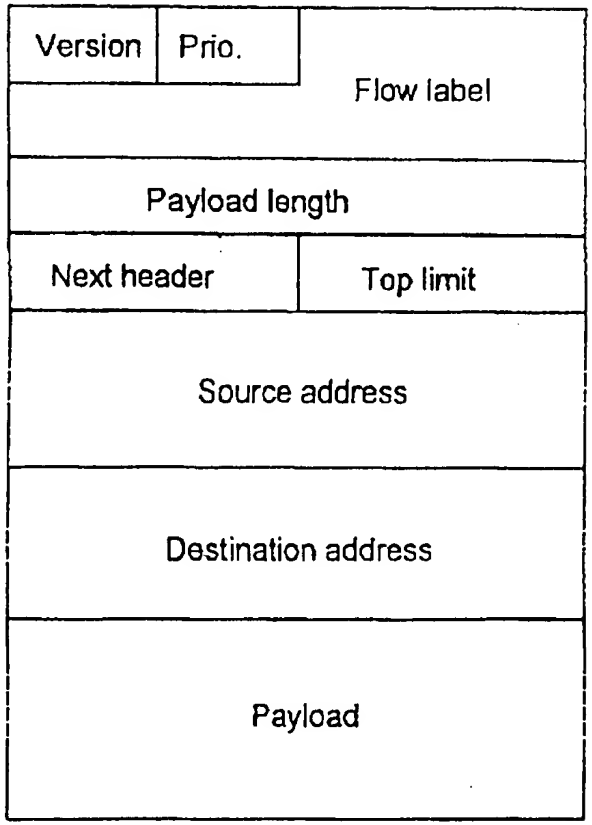

Version
Prio.
Flow label
Payload length
Next header
Top limit
Source address
Destination address
Payload

Fig. 2

This Page Blank (uspto)

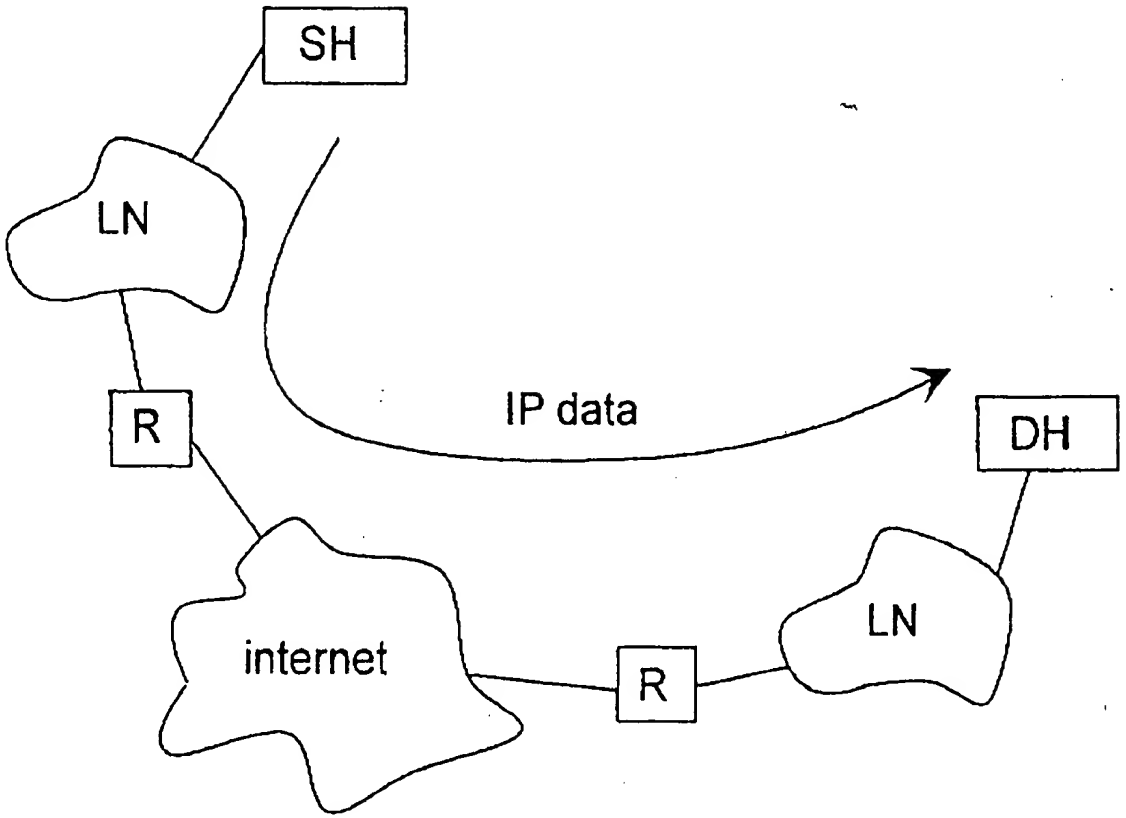


Fig. 3

This Page Blank (uspto)

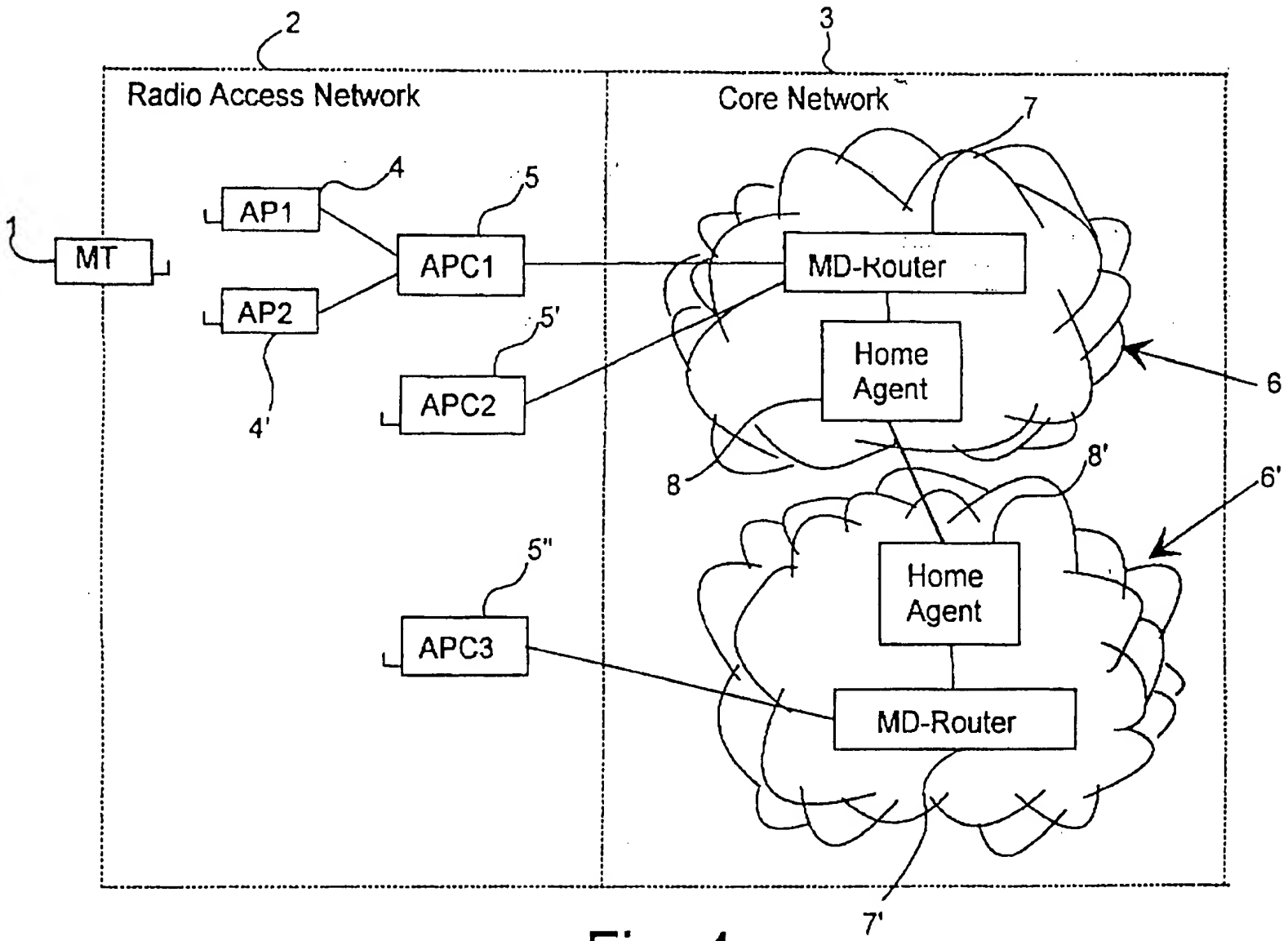

2
3
Radio Access Network
Core Network
7
4
5
AP1
1
MT
APC1
MD-Router
AP2
5'
Home
Agent
6
4'
APC2
8
8'
6'
5"
Home
Agent
APC3
MD-Router
7'

Fig. 4

This Page Blank (uspto)

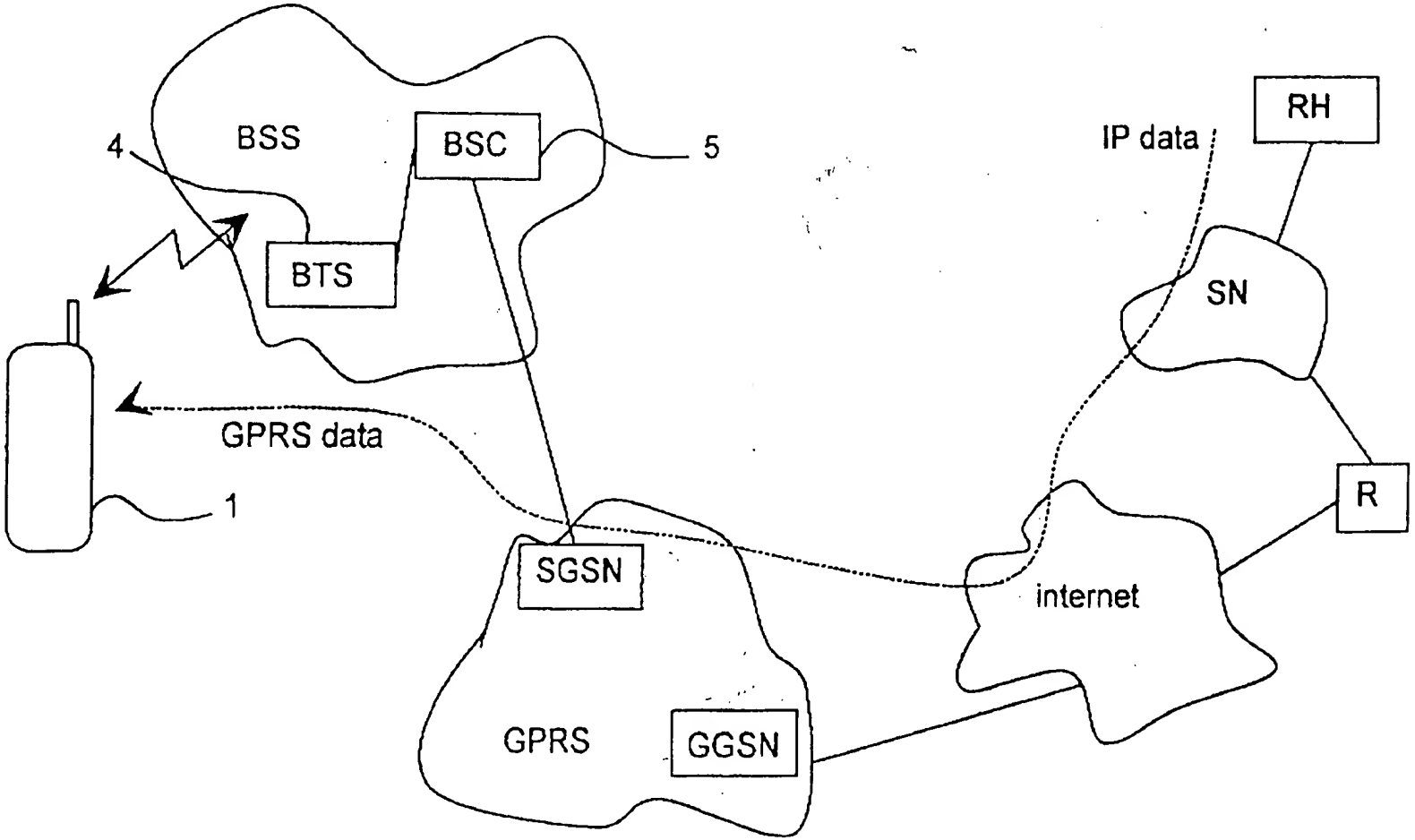

BSS
BSC
BTS
4
5
1
GPRS data
SGSN
GPRS
GGSN
IP data
RH
SN
R
internet

Fig. 5

This Page Blank (uspto)

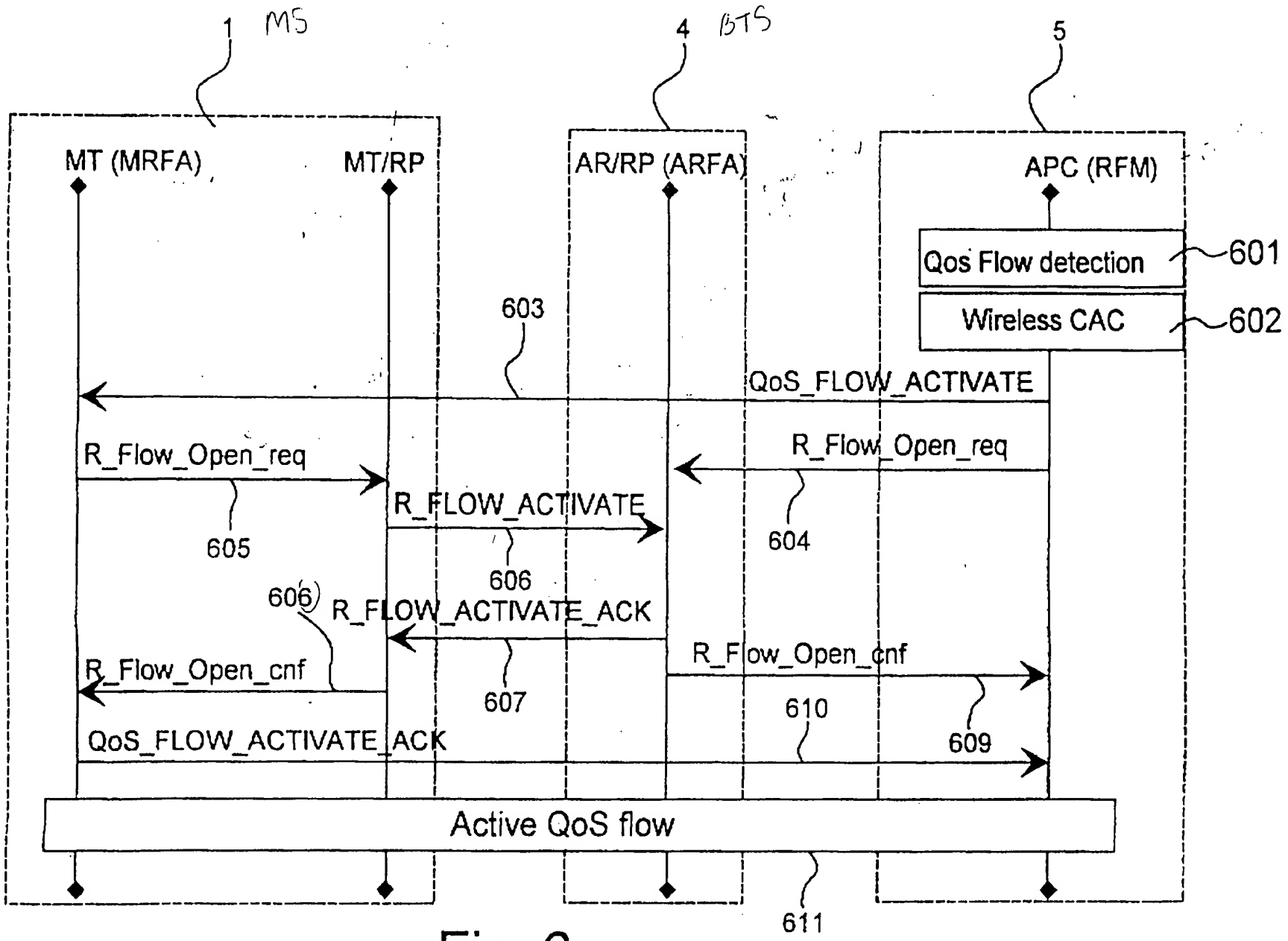

1
MS
4
BTS
5
MT (MRFA)
MT/RP
AR/RP (ARFA)
APC (RFM)
Qos Flow detection
601
Wireless CAC
602
603
QoS_FLOW_ACTIVATE
R_Flow_Open_req
R_Flow_Open_req
R_FLOW_ACTIVATE
605
604
606
606
R_FLOW_ACTIVATE_ACK
R_Flow_Open_cnf
R_Flow_Open_cnf
607
610
609
QoS_FLOW_ACTIVATE_ACK
Active QoS flow
611

Fig. 6

This Page Blank (uspto)

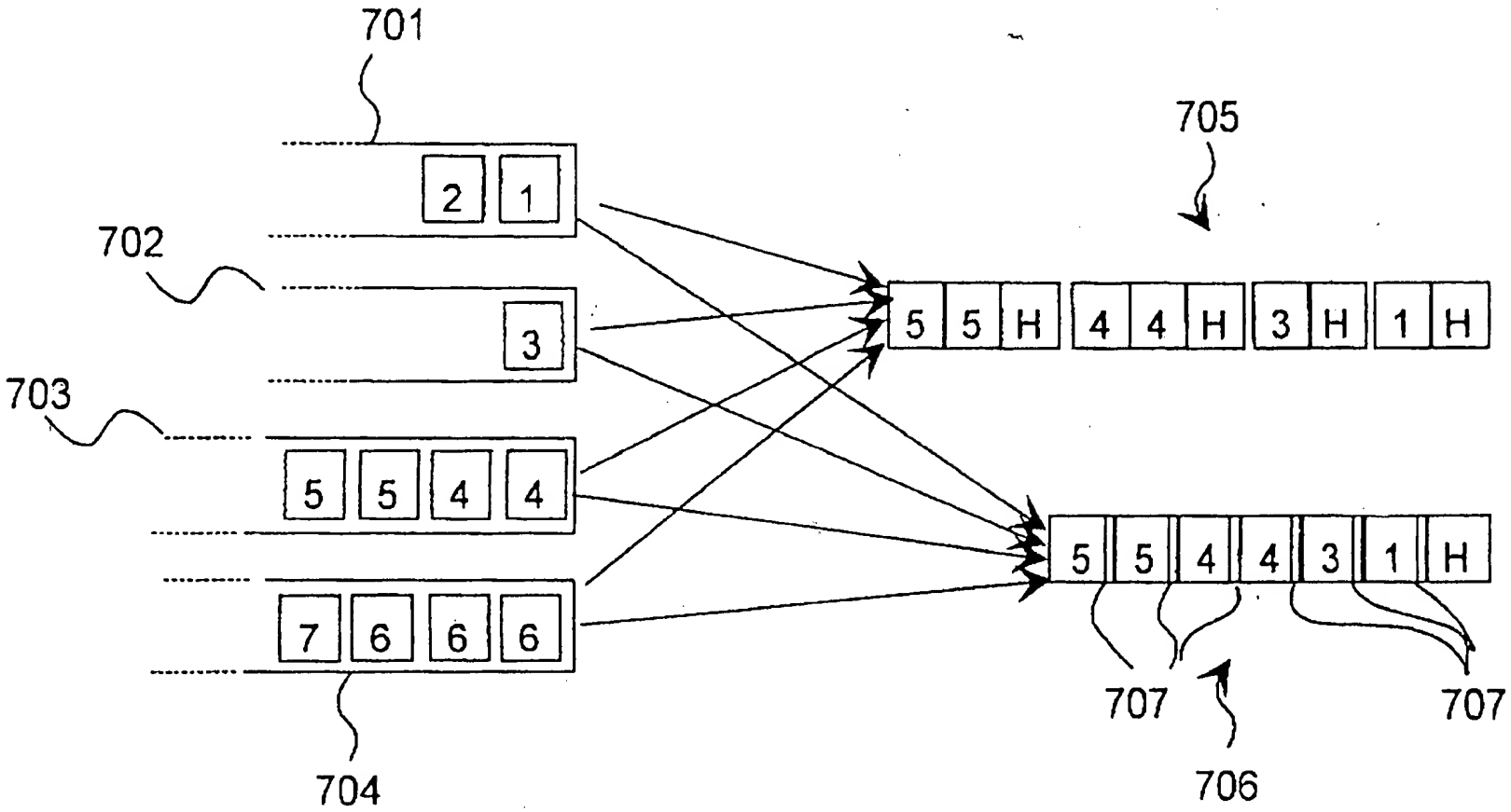

701
702
703
704
705
706
707
707
2
1
3
5
5
4
4
7
6
6
6
5
5
H
4
4
H
3
H
1
H
5
5
4
4
3
1
H

Fig. 7a

This Page Blank (uspto)

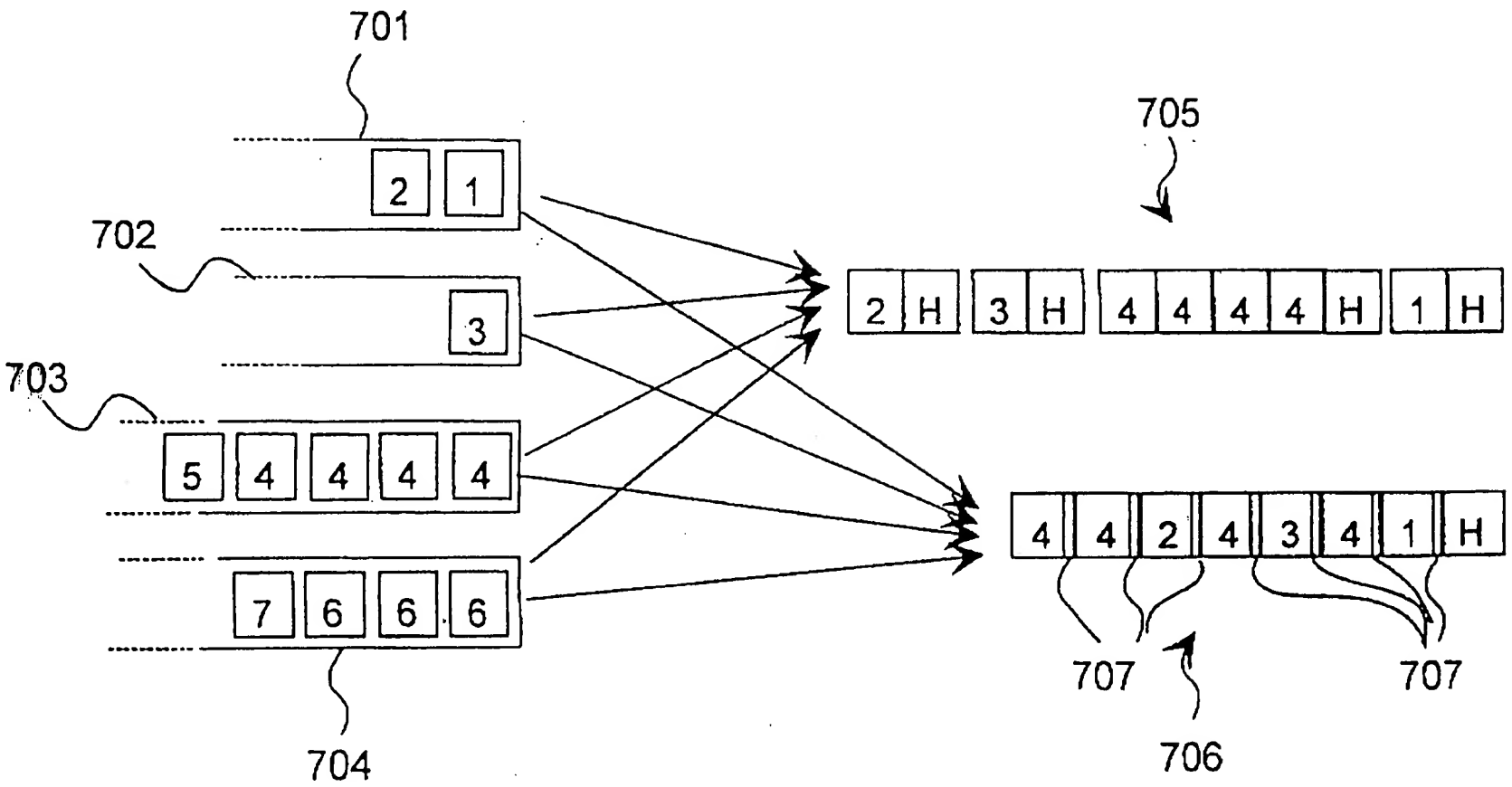

701
702
703
704
705
706
707
707
2
1
3
5
4
4
4
4
7
6
6
6
2
H
3
H
4
4
4
4
H
1
H
4
4
2
4
3
4
1
H

Fig. 7b

This Page Blank (uspto)

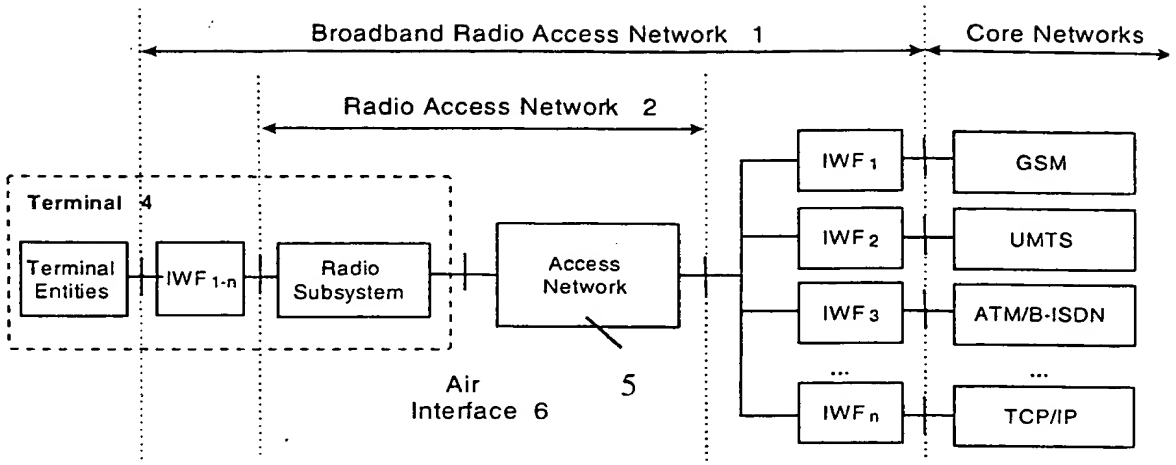


Figure 1

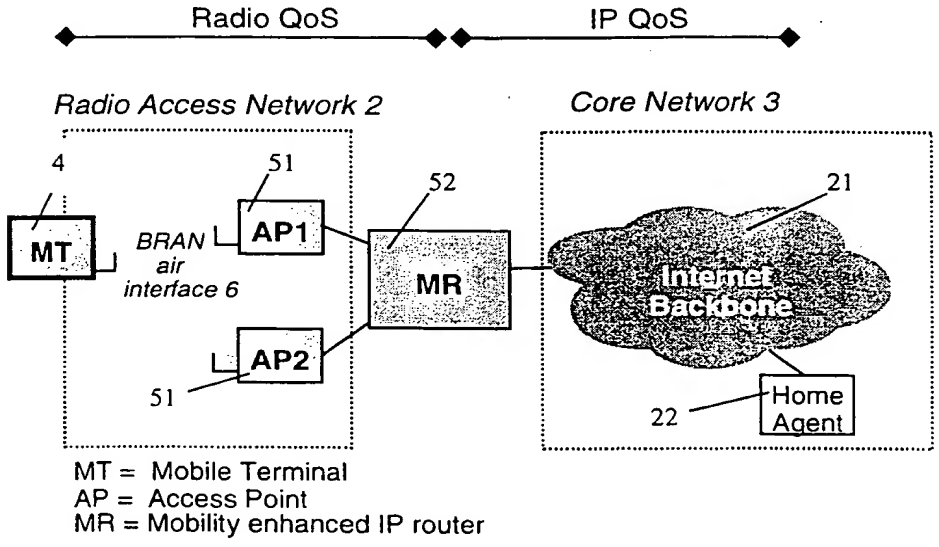


Figure 2

This Page Blank (uspto)

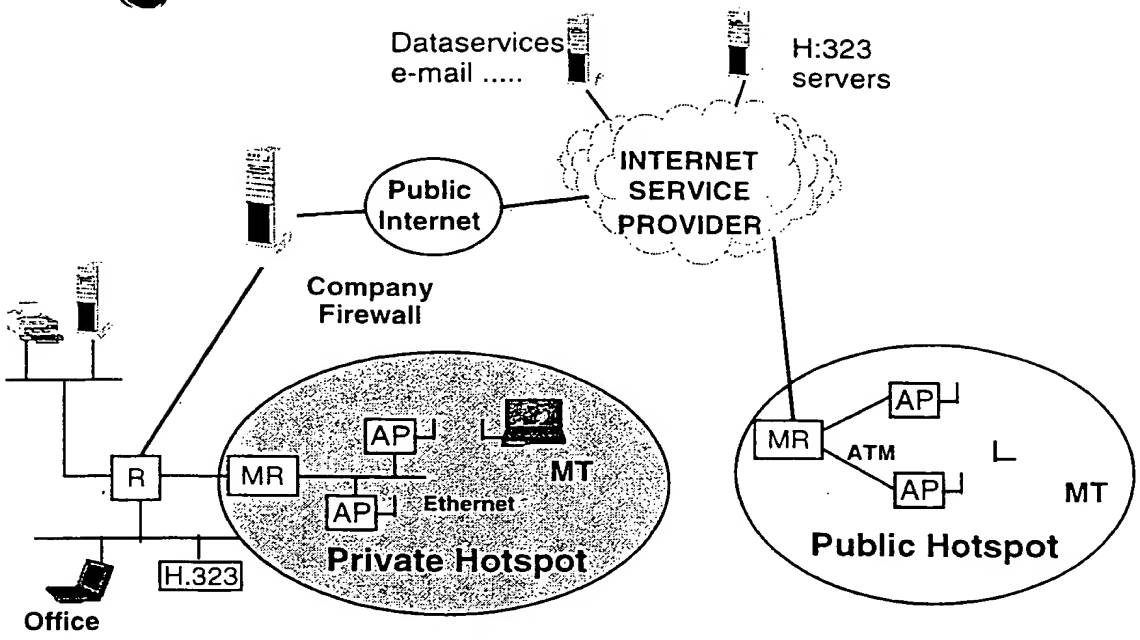


Figure 3

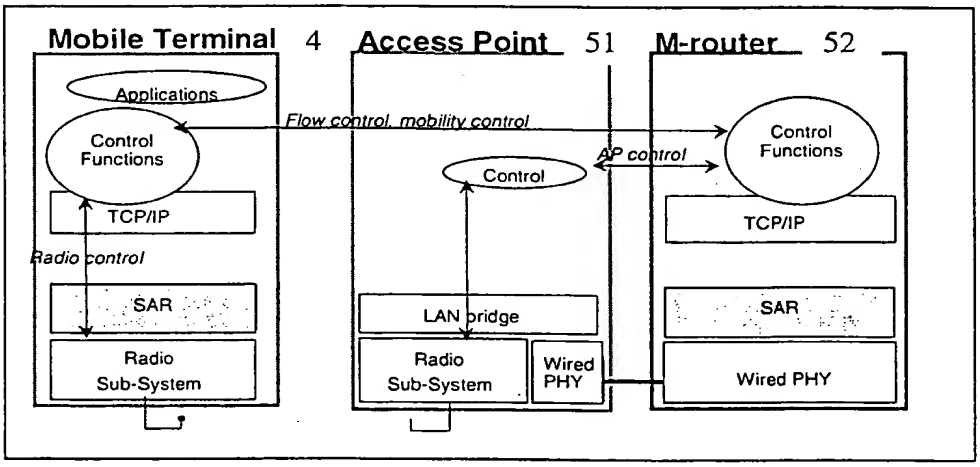
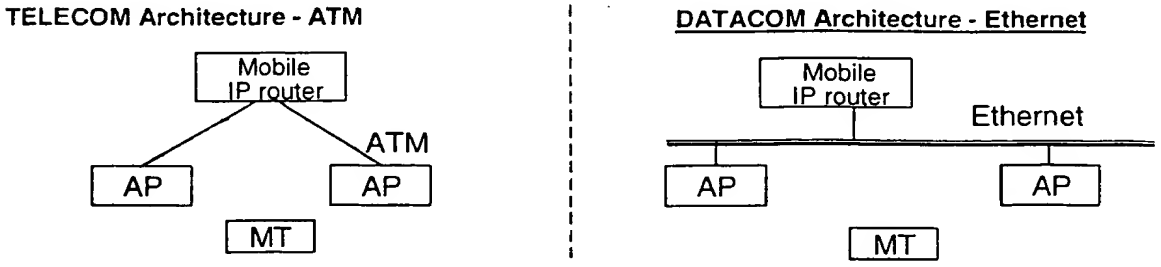


Figure 4



This Page Blank (uspto)

Figure 5

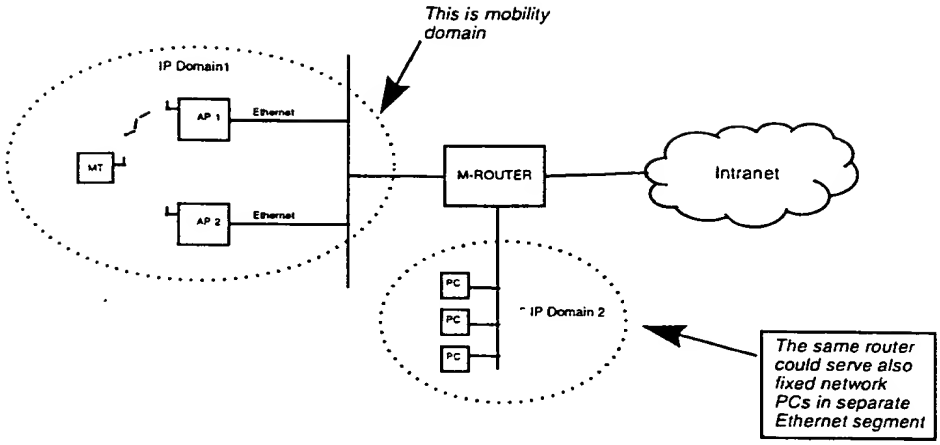


Figure 6

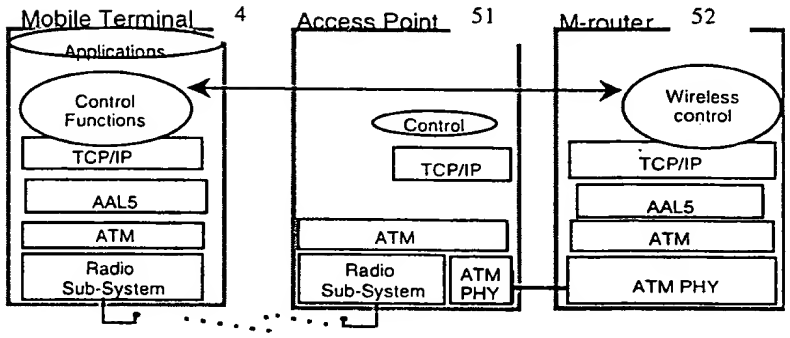


Figure 7

This Page Blank (uspto)

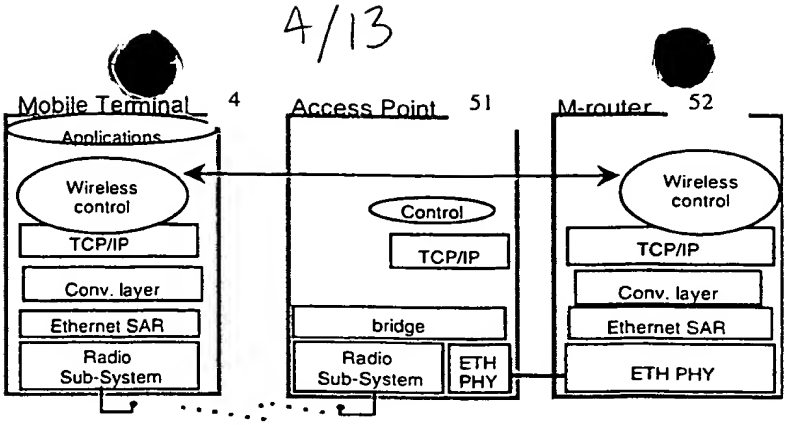


Figure 8

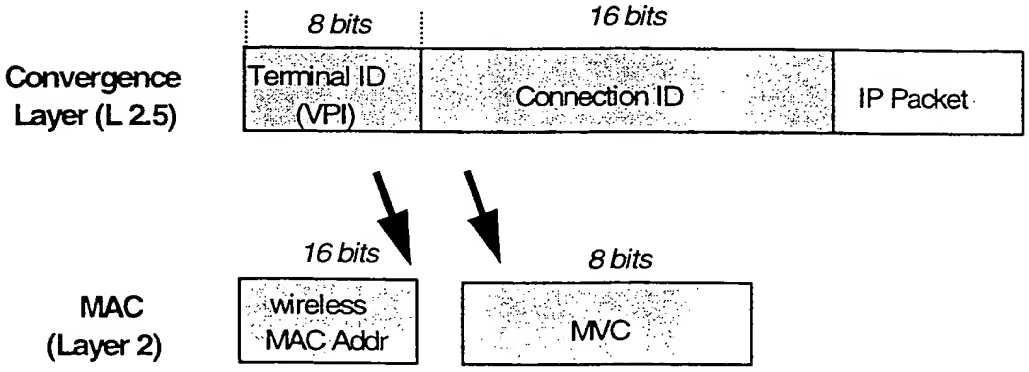


Figure 9

This Page Blank (uspto)

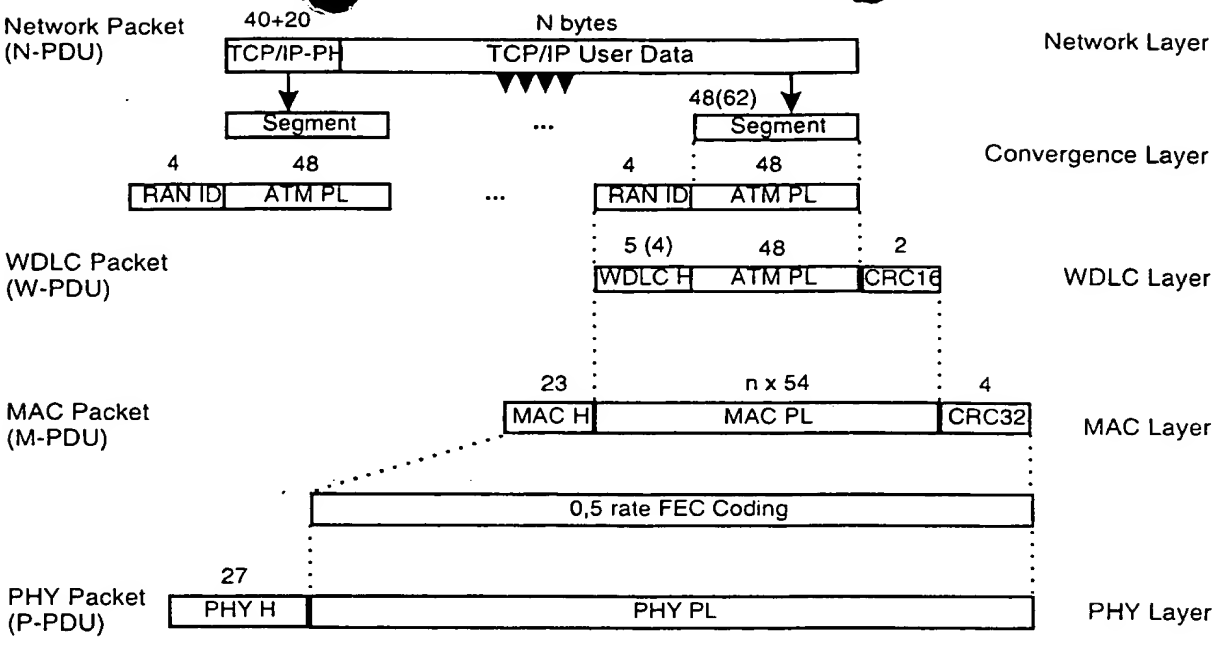


Figure 10

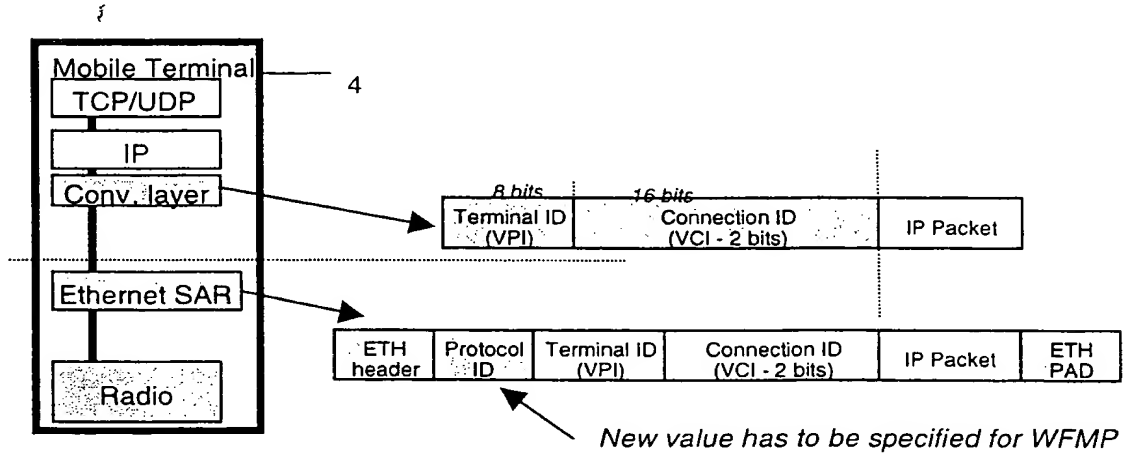


Figure 11

This Page Blank (uspto)

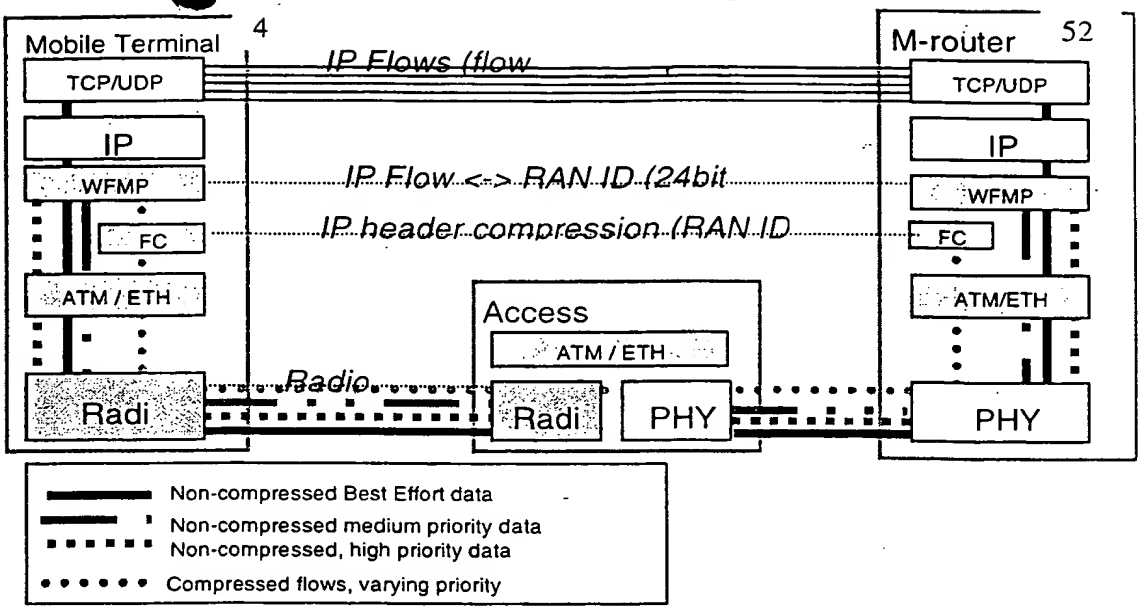


Figure 12

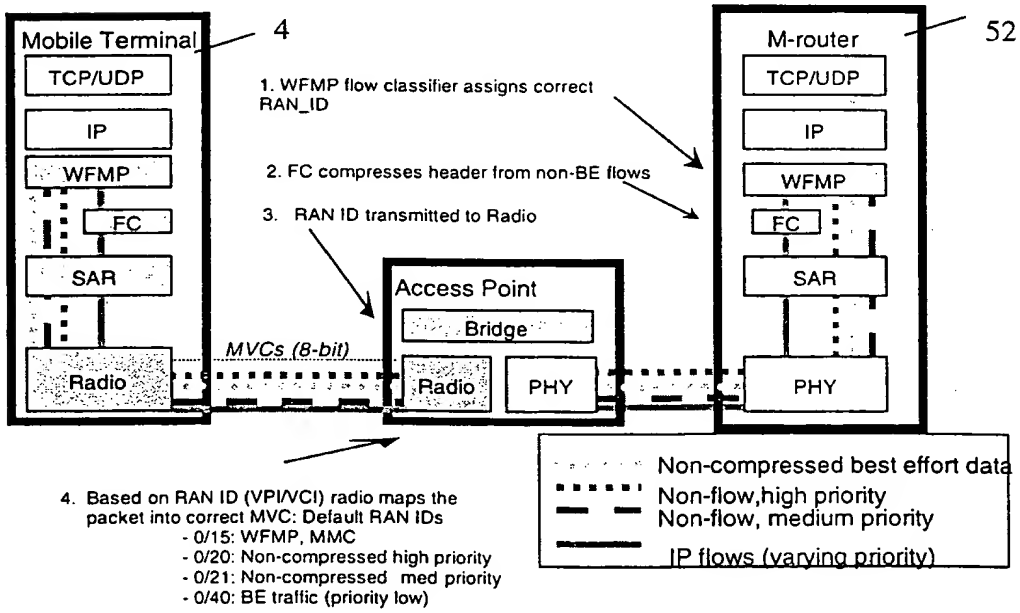


Figure 13

This Page Blank (uspto)

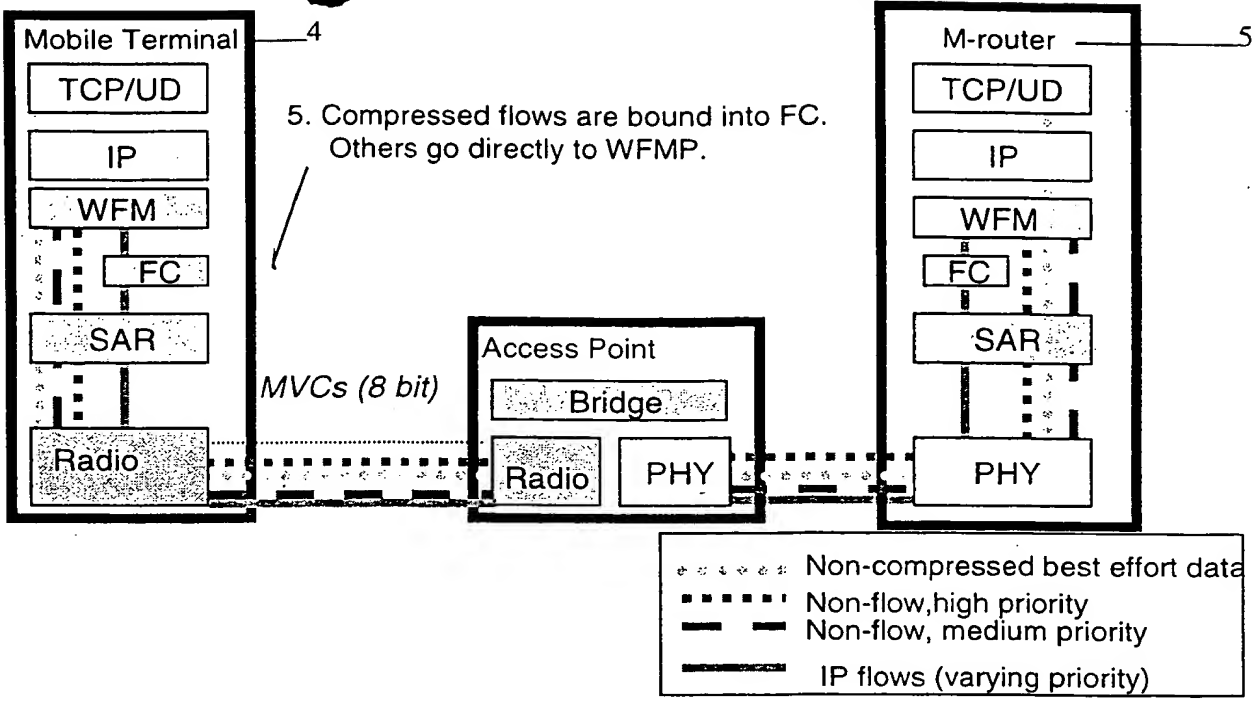


Figure 14

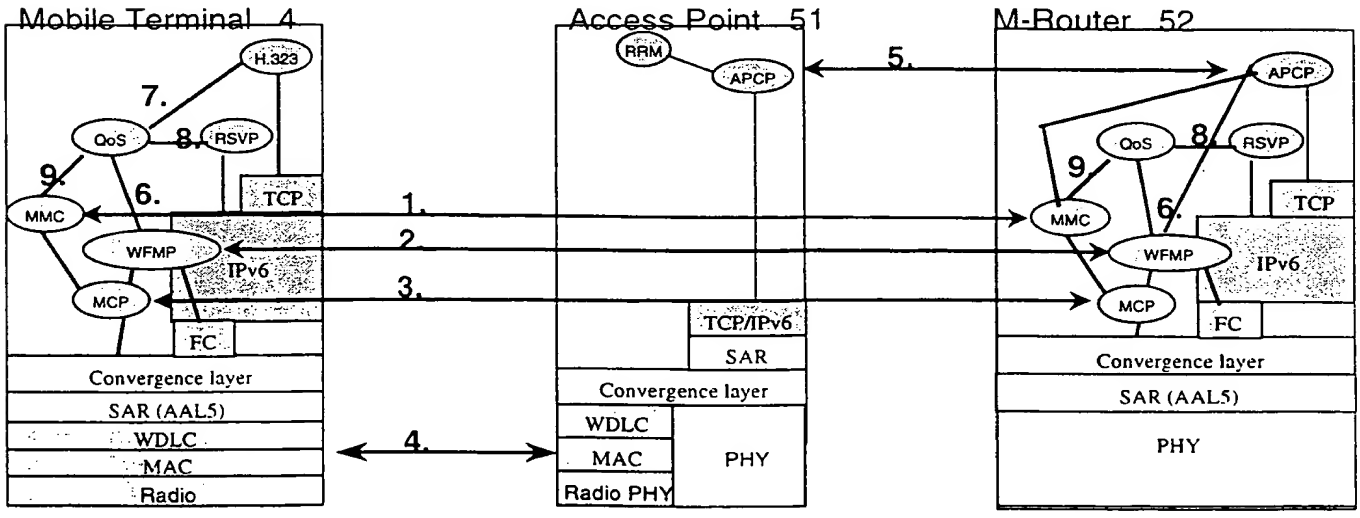


Figure 15

This Page Blank (uspto)

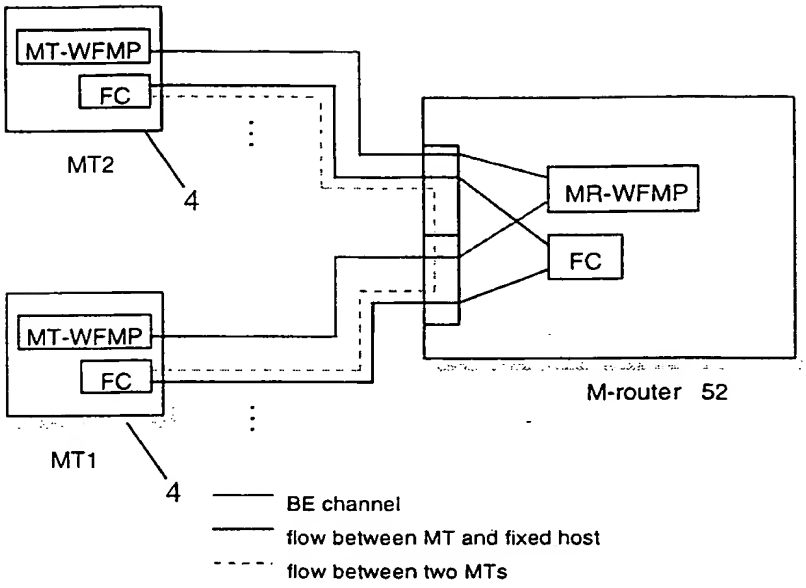


Figure 16

Active flows table

flow_type	dst_addr	src_addr	flow_label	protocol	dst_port	src_port	RAN_ID	AP_if

Default flows table

MT_id	RAN_ID_1	RAN_ID_2	RAN_ID_3	AP_if

Figure 17

This Page Blank (uspto)

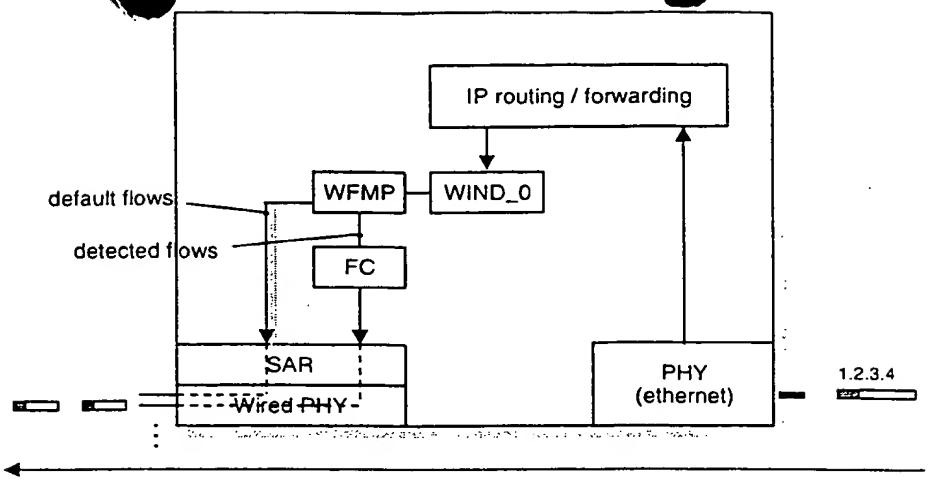


Figure 18

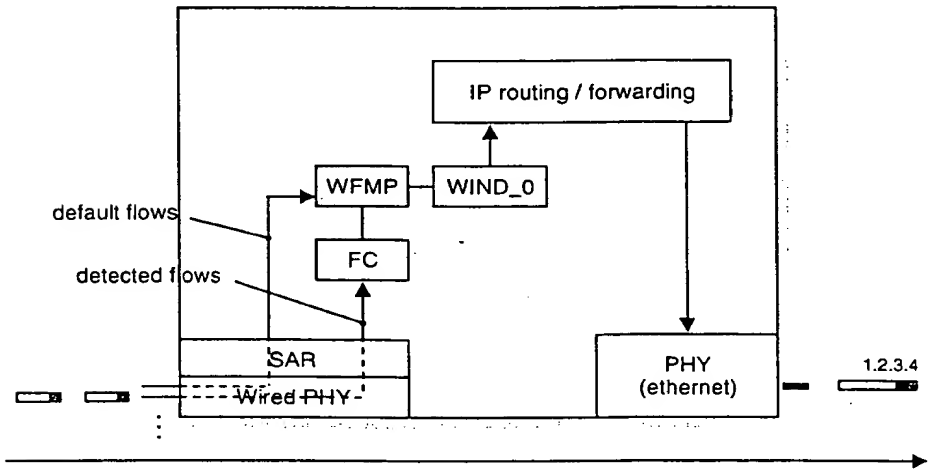


Figure 19

This Page Blank (uspto)

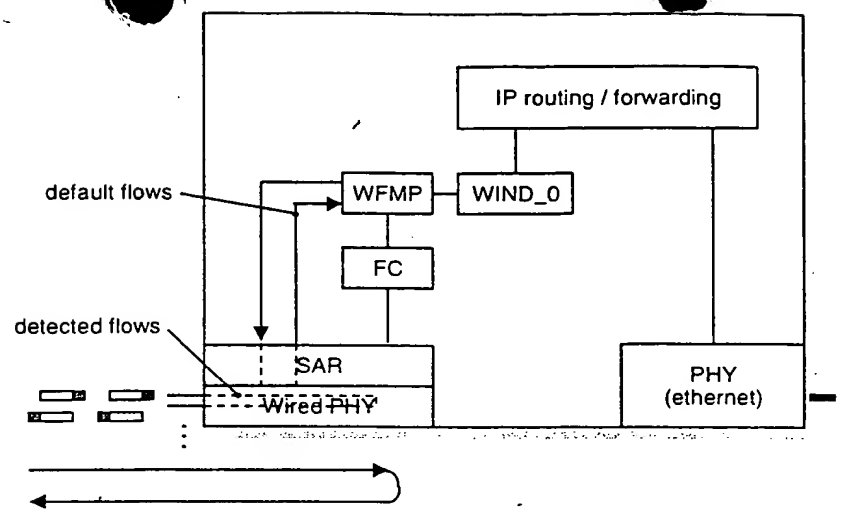


Figure 20

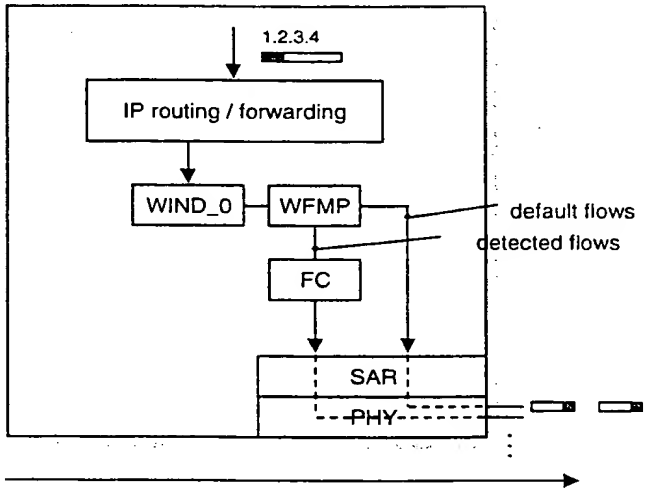


Figure 21

This Page Blank (uspto)

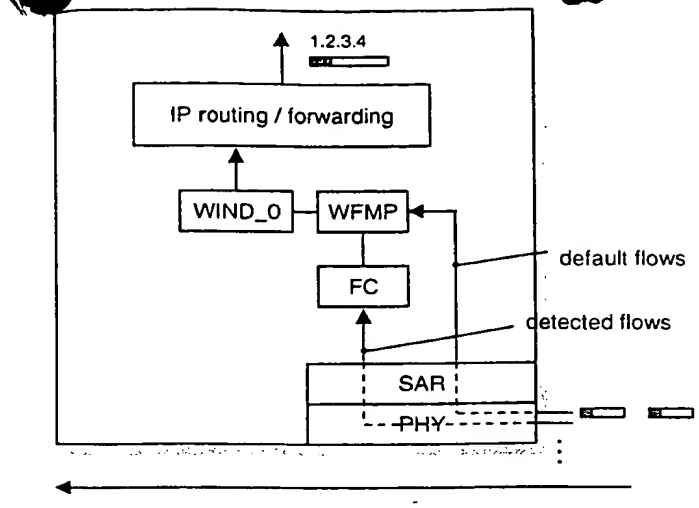


Figure 22

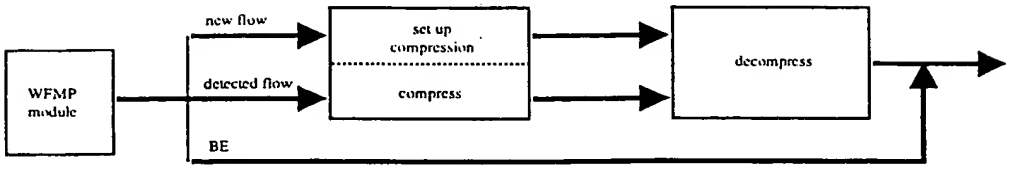


Figure 23

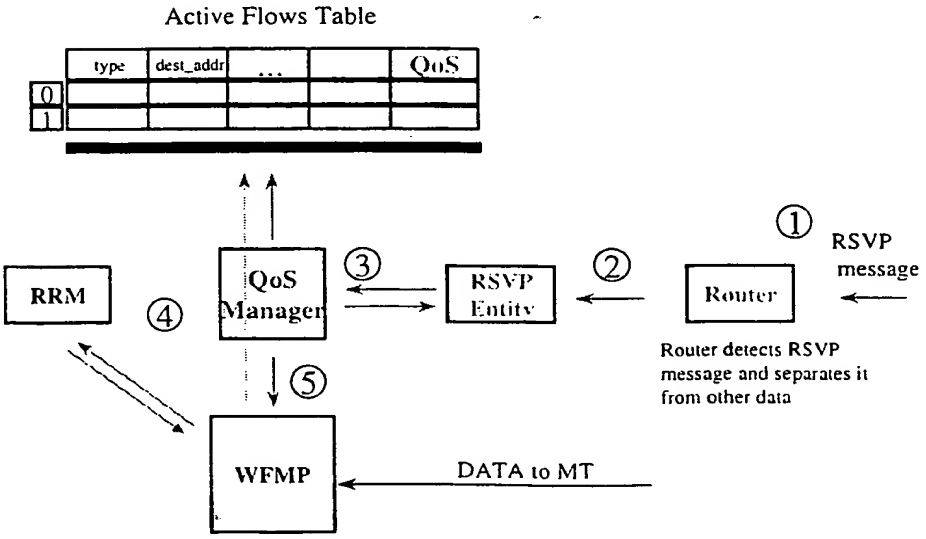


Figure 24

This Page Blank (uspto)

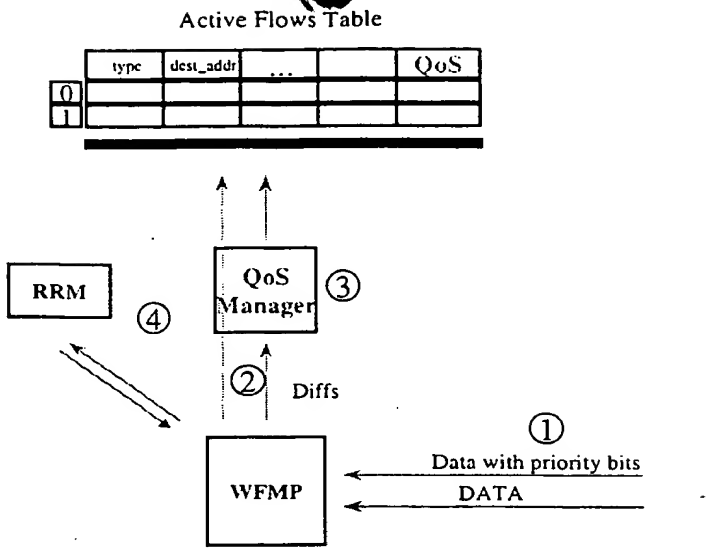


Figure 25

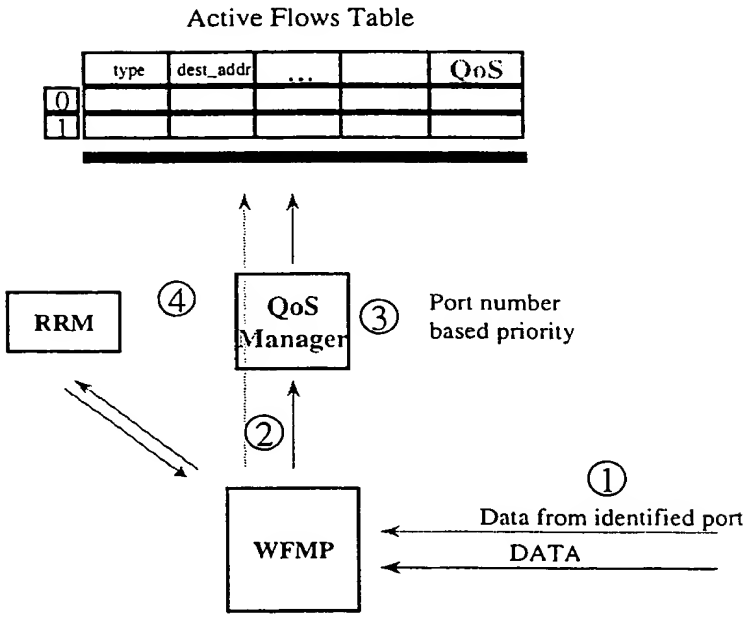


Figure 26

This Page Blank (uspto)

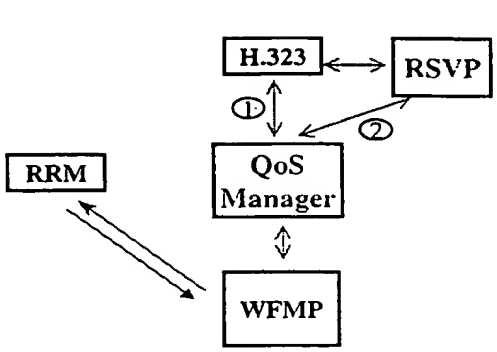


Figure 27

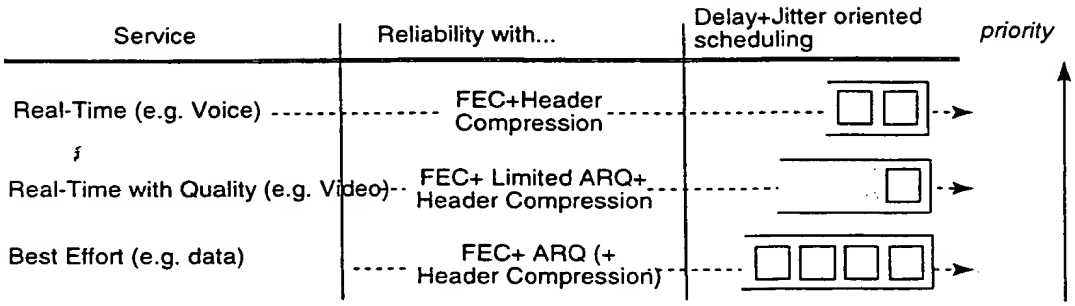


Figure 28

This Page Blank (uspto)

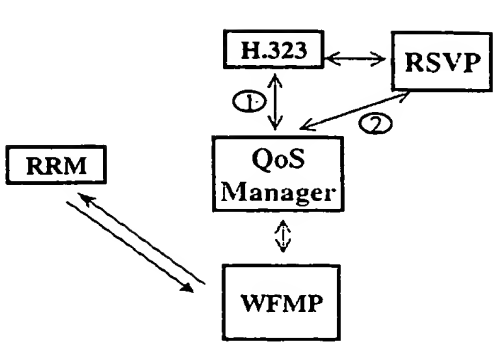


Figure 27

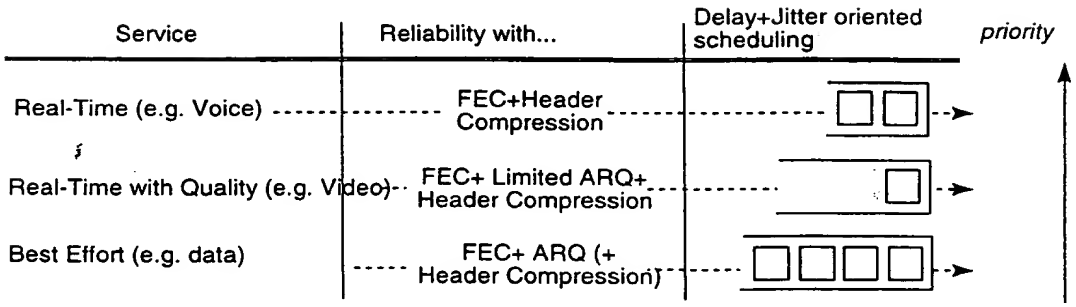


Figure 28

This Page Blank (uspto)

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.
As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

This Page Blank (uspto)